

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 30, 2004, 12:30:35 ; Search time 40.7195 Seconds
(without alignments)
2206.561 Million cell updates/sec

Title: US-10-021-723B-14

Perfect score: 1647
Sequence: 1 VSLLLAALLAALAPAA...CGSEAAACSLSDFAFLVAPAC 318

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

- 1: Geneseq1980s.*
- 2: Geneseq1990s.*
- 3: Geneseq2000s.*
- 4: Geneseq2001s.*
- 5: Geneseq2002s.*
- 6: Geneseq2003as.*
- 7: Geneseq2003bs.*
- 8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1647	100.0	318	5	ABP51936 Consensus
2	663.5	40.3	409	5	ABP51932
3	655.5	39.8	431	6	Aae37853 Rhizobium
4	655.5	39.8	437	6	Aae37854 pNOV4054
5	650	39.5	421	5	ABP51931
6	632.5	38.4	430	5	Aae15808 Escherich
7	627.5	38.4	430	6	Adal9451 E. coli K
8	627.5	38.1	436	7	AdC87743 Modified
9	616.5	37.4	412	7	Aae37851 Maize-opt
10	616.5	37.4	412	7	ABR82310 Nov9X phy
11	613.5	37.2	410	5	ABP51937 Phytase p
12	613.5	37.2	432	3	ABP51937 Phytase p
13	613.5	37.2	432	3	ABP51937 Phytase p
14	613.5	37.2	432	3	ABP51937 Phytase p
15	613.5	37.2	432	3	ABP51937 Phytase p
16	613.5	37.2	432	3	ABP51937 Phytase p
17	613.5	37.2	432	3	ABP51937 Phytase p
18	613.5	37.2	432	3	ABP51937 Phytase p
19	613.5	37.2	432	3	ABP51937 Phytase p
20	613.5	37.2	432	5	Aae15807 Escherich
21	613.5	37.2	432	5	ABP53625 Acid phos
22	613.5	37.2	432	6	Adal9450 E. coli K
23	613.5	37.2	432	6	ABU15427 Protein e
24	613.5	37.2	432	7	AdC87745 Escherich
25	611.5	37.1	432	4	Aae02631 E. coli a

26 611.5 37.1 440 4 AAB37892
27 611.5 37.1 440 5 AAE22836
28 611.5 37.1 440 5 AAE15806
29 611.5 37.1 440 6 ADA19446
30 611 37.1 476 5 ABP51933
31 610.5 37.1 432 4 AAE02632
32 610.5 37.1 432 4 AAE02635
33 608.5 36.9 432 3 AAY94753
34 608.5 36.9 432 6 ABR42161
35 605.5 36.8 432 4 AAE02634
36 602.5 36.6 432 6 ABR42162
37 596 36.2 441 5 ABP51934
38 590 35.8 441 6 ABU49868
39 589 35.8 441 5 ABP51929
40 567 34.4 420 5 ABP51930
41 556 33.8 423 2 AAY01513
42 304.5 18.5 261 4 ABG25698
43 203 12.3 123 4 AAU23269
44 141 8.6 393 5 ABP51935
45 138 8.4 426 4 ABG23348

ALIGNMENTS

RESULT 1
ABP51936
ID ABP51936 standard; protein; 318 AA.
XX
XX AC ABP51936;
XX
XX AC
DT 08-OCT-2002 (first entry)
XX
XX
DE Consensus phytase protein sequence SEQ ID NO:14.
XX
XX Yersinia pestis; phytase; enzyme; E.C.3.1.3.8; phytate; foodstuff; feed.
XX
XX Yersinia pestis.
XX
XX Synthetic.
XX
XX
PN WO200248332-A2.
XX
PD 20-JUN-2002.
XX
XX
PF 12-DEC-2001; 2001WO-US048774.
XX
XX
PR 12-DEC-2000; 2000US-0255090P.
XX
XX (DIVE-) DIVERSA CORP.
XX
XX Short J, Mathur BJ, Richardson T, Robertson D, Barton N;
XX
XX WPI; 2002-583504/62.
XX
XX Novel recombinant phytase protein and polynucleotide for improving
XX nutritional value of phytate-containing foodstuff, in animal feed and
XX feed supplements and to degrade excess phytase from environment or
XX sample.
XX
XX Claim 59; Fig 6; 208pp; English.

The present invention describes an isolated phytase protein (I). (I) can be used for improving the nutritional value of a phytate-containing foodstuff, by contacting the phytate-containing foodstuff with (I), where the enzyme catalyses the liberation of inorganic phosphate from the phytate-containing foodstuff, and so improving the nutritive value of the contacted foodstuff. The liberation of inorganic phosphate occurs prior to or after the ingestion of phytate-containing foodstuff by a recipient organism. Nucleotide sequences (II) encoding (I) can be used for producing an animal feed, by transforming a plant, plant portion or plant cell with a nucleic acid expression vector, comprising (II), culturing the plant, plant portion or plant cell under conditions in which the phytase protein is expressed and converting the plant, plant portion or


```

FH Key      Location/Qualifiers
FT Peptide  1..19
FT Protein  /label= Signal_peptide
FT          20..431
XX          /note= "Mature pNOV4054 phytase fusion protein"
FN WO2003057248-A1.
PD 17-JUL-2003.
XX
XX 30-DEC-2002; 2002WO-US041787.
XX
XX 28-DEC-2001; 2001US-0344476P.
XX
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
XX Lanahan MB, Betts S;
XX
XX WPI; 2003-607980/57.
XX
XX N-PSDB; AAD57148.
XX
XX Preparing a thermotolerant phytase for preparing animal feed or human
XX food by expressing in a plant cell an expression cassette comprising a
XX promoter operably linked to a nucleic acid molecule encoding a
XX thermotolerant phytase.
XX
XX Claim 63; Page 90; 157pp; English.
XX
XX The invention relates to a method for preparing a thermotolerant phytase.
XX The method comprises expressing in a plant cell an expression cassette
XX comprising a promoter operably linked to a nucleic acid molecule encoding
XX a thermotolerant phytase which retains at least 40% activity after 30
XX minutes at 60pluoc and has a specific activity of greater than 200 U/mg
XX at pH 4.5 and 37pluoc. The method is useful for preparing a
XX thermotolerant phytase for preparing animal feed or human food. The
XX invention is useful for reducing the feed conversion ratio and increasing
XX weight gain of animals fed diets with inorganic phosphate at levels below
XX 0.45%, minimising dietary requirements of phosphorus in an animal,
XX enhancing the utilisation of phosphorus present in animal feed, enhancing
XX organic phosphorus utilisation from organic phosphorus sources in animal
XX feed, decreasing the phosphate levels in excreta from an animal,
XX improving the processing of grain, improving the nutritive value of
XX processed grain product or a method of processing grain, improving the
XX nutritive value of animal feed and human food, and preparing a
XX transformed plant which expresses a thermotolerant phytase. The present
XX sequence is pNOV4054 phytase fusion protein used in the exemplification
XX of the invention
XX
XX Sequence 431 AA;
XX
XX Query Match      39.8%; Score 655.5; DB 6; Length 431;
XX Best Local Similarity 46.2%; Pred. No. 6.2e-56;
XX Matches 198; Conservative 43; Mismatches 75; Indels 113; Gaps 34;
XX
XX QY 1 VSLLAALLAALAPAAAAP-LKLEKVVILSRHGRVSPTKA--LM--VSPD-WP-WP 53
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX Db 3 VLLVALALLAAGATSAAQSEPELKESVIVSRHGVRAPTKATQLMQDVTPDWPTW 62
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX QY 54 V--GLTPRGALV-LIQT-----RYFARGILP--CCPAAG--TILADVDERTR-TGQAPA 100
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX Db 63 VKLGELTRGELIAYLGHYWRQLVADGLLPKCGCPGSGQVAILADVDERTRTKTGEAFA 122
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX QY 101 AGLAPGCAI-VH-----AGDDPIIFHGLDTG-C-LDQA---DAILAAGEGGITARHG---- 145
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX Db 123 AGLAPDCAITVHTQADTSSPDLEPLNPKLTKGVQCQLDNANVTDAILLERAGGSIADPTGHYQT 182
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX QY 146 --LTIKVLNF--ASACI-----ECARVVVRIGLPILLASTLSE 178
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX Db 183 AFRLERLVLPQNLCKRKQDESCSLTQALPSELKVKSGDCVSLTGAVS--LASMLTE 240
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX QY 179 IFLLAYAGCPMEVGWGRV--SAEW----LIHNAQF-LMNRTPYIA--RAYPI--LIIVTAL 227
XX      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

241 IFLLQQAQMPGPGWGRITDSHQWNTLLSLHNAQFDLLQRTPEVARSRATPLLDLTKTAL 300

228 S-----PARVVLLAGHTNLA-LGG-LDL-WQLP-QDETTPGALVFEIWN 270

301 TPHPQKQAYGVTLTSLVFIAGHDTNLANLGCALELANWTLPGQPDNTPPGGELYFERWR 360

271 R-----YRVVMYQTMQDLRNLEPLLPRIPL-----IPGC---GSEAAACSLSDFARL 313

361 RLSDNSQMTQVSLVFQTLQOMEDKTPLSINTPPGGEVKLTLAGCEERNAQCMCSLAGFTQI 420

314 V-----APAC 318

421 VNEARIPAC 429

RESULT 4

AAE37854

ID AAE37854 standard; protein; 437 AA.

XX AC AAE37854;

XX DT 06-NOV-2003 (first entry)

XX DE pNOV4058 phytase fusion protein.

XX KW Thermotolerant phytase; weight gain; animals fed diet; grain processing;

XX KW nutritive value; transformed plant; anabolic; chimeric; maize.

XX OS Chimeric - Zea mays.

XX OS Chimeric - Unidentified.

XX FH Key Location/Qualifiers

FT Peptide 1..19

FT Protein /label= Signal_peptide

FT Protein 20..437

FT Region /note= "Mature pNOV4058 phytase fusion protein"

FT Region 432..437

FT /note= "SEKDEL signal"

XX WO2003057248-A1.

PN 17-JUL-2003.

XX 30-DEC-2002; 2002WO-US041787.

XX 28-DEC-2001; 2001US-0344476P.

XX (SYGN) SYNGENTA PARTICIPATIONS AG.

XX Lanahan MB, Betts S;

XX WPI; 2003-607980/57.

XX N-PSDB; AAD57149.

XX Preparing a thermotolerant phytase for preparing animal feed or human

XX food by expressing in a plant cell an expression cassette comprising a

XX promoter operably linked to a nucleic acid molecule encoding a

XX thermotolerant phytase.

XX Example 1; Page 91; 157pp; English.

XX The invention relates to a method for preparing a thermotolerant phytase.

XX The method comprises expressing in a plant cell an expression cassette

XX comprising a promoter operably linked to a nucleic acid molecule encoding

XX a thermotolerant phytase which retains at least 40% activity after 30

XX minutes at 60pluoc and has a specific activity of greater than 200 U/mg

XX at pH 4.5 and 37pluoc. The method is useful for preparing a

XX thermotolerant phytase for preparing animal feed or human food. The

XX invention is useful for reducing the feed conversion ratio and increasing

XX weight gain of animals fed diets with inorganic phosphate at levels below

XX 0.45%, minimising dietary requirements of phosphorus in an animal, enhancing

XX the utilisation of phosphorus present in animal feed, enhancing organic

XX phosphorus utilisation from organic phosphorus sources in animal feed,

XX decreasing the phosphate levels in excreta from an animal, improving

XX the processing of grain, improving the nutritive value of processed

XX grain product or a method of processing grain, improving the nutritive

XX value of animal feed and human food, and preparing a transformed plant

XX which expresses a thermotolerant phytase. The present sequence is

XX pNOV4054 phytase fusion protein used in the exemplification of the

XX invention

XX Sequence 431 AA;

XX Query Match 39.8%; Score 655.5; DB 6; Length 431;

XX Best Local Similarity 46.2%; Pred. No. 6.2e-56;

XX Matches 198; Conservative 43; Mismatches 75; Indels 113; Gaps 34;

XX QY 1 VSLLAALLAALAPAAAAP-LKLEKVVILSRHGRVSPTKA--LM--VSPD-WP-WP 53

XX :

XX Db 3 VLLVALALLAAGATSAAQSEPELKESVIVSRHGVRAPTKATQLMQDVTPDWPTW 62

XX :

XX QY 54 V--GLTPRGALV-LIQT-----RYFARGILP--CCPAAG--TILADVDERTR-TGQAPA 100

XX :

XX Db 63 VKLGELTRGELIAYLGHYWRQLVADGLLPKCGCPGSGQVAILADVDERTRTKTGEAFA 122

XX :

XX QY 101 AGLAPGCAI-VH-----AGDDPIIFHGLDTG-C-LDQA---DAILAAGEGGITARHG---- 145

XX :

XX Db 123 AGLAPDCAITVHTQADTSSPDLEPLNPKLTKGVQCQLDNANVTDAILLERAGGSIADPTGHYQT 182

XX :

XX QY 146 --LTIKVLNF--ASACI-----ECARVVVRIGLPILLASTLSE 178

XX :

XX Db 183 AFRLERLVLPQNLCKRKQDESCSLTQALPSELKVKSGDCVSLTGAVS--LASMLTE 240

XX :

XX QY 179 IFLLAYAGCPMEVGWGRV--SAEW----LIHNAQF-LMNRTPYIA--RAYPI--LIIVTAL 227

XX :


```

XX DE Escherichia coli appA phytase mutant protein.
XX KW Bacterial phytase; K12 appA phytase; protease stability; anabolic;
XX KW gastrointestinal; nutritional value; feed treatment process; therapy;
XX KW thermal tolerance; growth performance; alcoholic drink; biopulping;
XX KW non-alcoholic drink; biobleaching; mutant; muten.
XX OS Escherichia coli.
XX OS Synthetic.
XX FH Key Location/Qualifiers
FT FT Misc-difference 68 /note= "Wild type Trp substituted with Glu"
FT FT Misc-difference 84 /note= "Wild type Gln substituted with Trp"
FT FT Misc-difference 95 /note= "Wild type Ala substituted with Pro"
FT FT Misc-difference 97 /note= "Wild type Lys substituted with Cys"
FT FT Misc-difference 168 /note= "Wild type Ser substituted with Glu"
FT FT Misc-difference 180 /note= "Wild type Arg substituted with Tyr"
FT FT Misc-difference 225 /note= "Wild type Asn substituted with Cys"
FT FT Misc-difference 276 /note= "Wild type Tyr substituted with Asp"
XX KW WO200190333-A2.
XX KW 29-NOV-2001.
XX KW 24-MAY-2001; 2001WO-US017118.
XX KW 25-MAY-2000; 2000US-00580515.
XX KW (DIVE-) DIVERSA CORP.
XX KW Short JM, Kretz KA, Gray KA, Barton NR, Garrett JB, O'donoghue E;
XX KW WPI; 2002-083108/11.
XX KW New bacterial phytase for e.g. improving the nutritional value of phytate
XX KW -containing foodstuffs and subsequently improving the growth performance
XX KW of an organism that consumes it, or in treating animal digestive systems.
XX KW Claim 59; Fig 8; 170pp; English.
XX KW The patent discloses recombinant bacterial phytase from Escherichia coli
XX KW K12 appA phytase. The enzyme has phytase activity and improved thermal
XX KW tolerance when compared with wild-type phytase. It has improved protease
XX KW stability at low pH. The recombinant phytase is useful for improving the
XX KW nutritional value of phytate-containing foodstuffs and subsequently
XX KW improving the growth performance of an organism that consumes it, in
XX KW treating animal digestive systems, in feed treatment processes and for in
XX KW vitro purposes related to research, discovery and development. They are
XX KW also used for generating recombinant digestive system life forms, for
XX KW producing or manufacturing alcoholic and non-alcoholic drinks based on
XX KW the use of moulds, grains and/or plants, in biopulping and bio-bleaching
XX KW where a reduction in the use of environmentally harmful chemicals that
XX KW are traditionally used in the pulp and paper industry is desired and in
XX KW the reduction or possible elimination of the need for mineral
XX KW supplements, enzymes or therapeutic drugs for animals from the daily feed
XX KW thus increasing the amount calories and nutrients present in the feed.
XX KW The present sequence is E. coli appA phytase mutant protein
XX KW Sequence 430 AA;
SQ Query Match 38.4%; Score 632.5; DB 5; Length 430;
Best Local Similarity 45.2%; Pred. No. 1.2e-53;
Matches 194; Conservative 44; Mismatches 76; Indels 115; Gaps 35;

```

```

QY 2 SLLLAALLAALAP--AAAAAP-LKLEKVVILSRHGRSPTKA--LM--VSPD-WP-W 52
DB 3 AILPFLSLIPLTPQSAFAQSEPELKLESVIVSRHGRVAPTKATQLMQDVTPDANPTW 62
QY 53 PV---GLTPRGAALV-LIGY----RYFARGLLP--GCPAAG--TILADVDERTR--TGOAF 99
DB 63 PVKLGELTPRGELIAYLGHYWRQLVADGLLPCGCPQSGQVAILADVDERTRKTGEAF 122
QY 100 AAGLAPGCAI-VH-----AGDDPIFHGLDTG-C-LDQR---DAILAAGEGLTARHG--- 145
DB 123 AAGLAPDCAITVHTQADTSSPDPLFNPLKTVGCQLDNANVTDAILEAG--GSIADPTGHYQ 181
QY 146 ---LTLAKLVNF--ASACL-----ECARVVVGRIGPLIASTLS 177
DB 182 TAPRELRLVNFPPQSNLCKREKQDECSLTQALPSELKVSADCVSLTGAVS--LASMLT 239
QY 178 EIFLLEYAQGMVEVCGRI--SAEW-----LHNAQFLMNRTPYIA--RATPI--LIVTAL 227
DB 240 EIFLLQQAQGMPEPCGWRITDSHOWNTLLSLHNAQFDLQRTPEVARSATPLDLLDKTAL 299
QY 228 S-----PARVLLAGHDNTLA-LGG-LDL-WQLP-OPDETPPGCALVFELWN 270
DB 300 TPHPKQKQAYGVTLPTSVLPFIAGHDNTLANLGGALSLNWTLPQGDDNTPPGGELVFERWR 359
QY 271 R-----YVRVMYQTMQDLRNLEPLPRILP-----IPGC---GSEAAACSLSDPARL 313
DB 360 RLSDNSQWISLVFQTLQQRDKTPLSLNTPPGREVKLTLAGCBERNAGCMCSLAGTQI 419
QY 314 V---APAC 318
DB 420 VNEARIPAC 428

```

RESULT 7

ADA19451

ID ADA19451 standard; protein; 430 AA.

XX ADA19451;

XX ADA19451;

XX 20-NOV-2003 (first entry)

XX E. coli K12 phytase mutant.

XX Phytase; enzyme; phytate; appA gene; animal feed; inorganic phosphate;

XX KW digestion enhancement; transgenic; thermal tolerance; protease stability;

XX KW mutant; muten.

XX OS Synthetic.

OS Escherichia coli; strain K12.

XX Key Location/Qualifiers

FH Misc-difference 68 /note= "Wild-type Trp substituted by Glu"

FT Misc-difference 72 /note= "Encoded by CGN"

FT Misc-difference 84 /note= "Wild-type Gln substituted by Trp"

FT Misc-difference 95 /note= "Wild-type Ala substituted by Pro"

FT Misc-difference 97 /note= "Wild-type Lys substituted by Cys"

FT Misc-difference 168 /note= "Wild-type Ser substituted by Glu"

FT Misc-difference 180 /note= "Wild-type Arg substituted by Tyr"

FT Misc-difference 226 /note= "Wild-type Asn substituted by Cys"

FT Misc-difference 277 /note= "Wild-type Tyr substituted by Asp"

XX US2002136754-A1.

XX 26-SEP-2002.

Db 300 LFIAGHDTNLALGGALNLTLPQDPNTPPGGELVFERWRRLSDNSQWQVSLVFQTL 359
Qy 281 DQRLNLEPLPRILP-----IPGC---GSEACSLSDPARLV---APAC 318
Db 360 QQVRDRTPLSLNTPPGGVKLTLAGCERNAQGWCSLAGFTQIVNEARIPAC 410

RESULT 10

ABR82310
ID ABR82310 standard; protein; 412 AA.

XX ABR82310;
XX 06-NOV-2003 (first entry)

XX Nov9X phytase polypeptide.

XX Nov9X; thermotolerant; phytase; phosphate; animal food; phosphorous;
KW enzyme.
XX Synthetic.

XX Key Location/Qualifiers
FH Misc-difference 1 /note= "encoded by GAG"
FT W02003057247-A1.
PN 17-JUL-2003.

XX 30-DEC-2002; 2002WO-US041785.
XX 28-DEC-2001; 2001US-0344523P.

XX (SYGN) SYNGENTA PARTICIPATIONS AG.
XX (DIVE-) DIVERSA CORP.
XX Lanahan ML, Koepf E, Kretz K;
XX WPI; 2003-598328/56.
XX N-PSDB; ACF45789.

XX Preparing thermotolerant phytase useful for preparing animal feed or
human food, by expressing in microbial host cell expression cassette
comprising a promoter operably linked to a nucleic acid encoding the
enzyme.
XX Claim 11; Fig 1B; 79pp; English.

XX The invention relates to preparing a thermotolerant phytase (I),
involving expressing in a microbial host cell an expression cassette
comprising a promoter operably linked to a nucleic acid molecule encoding
(I). The thermotolerant phytase has a specific activity of greater than
400 U/mg, 600 U/mg or 800 U/mg at pH 4.5 and 37plusoc. The nucleic acid
molecule encodes a fusion polypeptide comprising the thermotolerant
phytase and a signal sequence which is operably linked to the
thermotolerant phytase. The thermotolerant phytase is glycosylated and
has a simulated gastric half-life of greater than 25 minutes at a pH
greater than 2.0 and less than 4.0. (I) is useful for preparing animal
feed, improving nutritive value of animal feed and human food, and for
preparing human food. A heat-treated animal feed mixture comprising an
inorganic phosphate at below 0.45% and (I) is useful for reducing feed
conversion ratios or increasing weight gain of animals fed diets with
inorganic phosphate at levels below 0.45%. It is also useful for
decreasing phosphate levels in excreta from an animal. An animal feed
composition comprising (I) is useful for decreasing the feed conversion
ratio and increasing the weight gain of an animal. The composition is
useful for minimizing inorganic dietary requirements of phosphorous in an
animal, for enhancing utilization of phosphorous present in feed for an
animal, and for increasing organic phosphate utilization from organic
phosphorus sources in feed for an animal, where the feed increases the
bioavailability of inorganic phosphorus in the feed to the animal. It is

CC also useful for decreasing phosphate levels in excreta from an animal. In
CC all the above mentioned conditions, the feed is a poultry or swine feed.
CC The thermotolerant phytase has a half-life of about 30 minutes in the
CC digestive tract of the animal. The present sequence represents a
CC thermotolerant Nov9X phytase
XX Sequence 412 AA;

Query Match 37.4%; Score 616.5; DB 7; Length 412;

Best Local Similarity 45.5%; Pred. No. 4.2e-52;
Matches 187; Conservative 41; Mismatches 70; Indels 113; Gaps 34;

Qy 19 AAAP-LKLEKWLILSRHGVSRPTKA--LM--YSPD-WP-WPV---GLTPRGALV-LLG 67
Db 2 AQSEPELKLESVIVSRHGVRAPTKATQLMQDTPDAMPVFKLGLTTRGGELIAYLG 61

Qy 68 Y-----RYFARGLLP--CCPAAG--TILADVDETR-TGOAFAAGLAPGCAI-VH-----A 112
Db 62 HYWRQLVADGLLPKCGCQSGQVAITADYVDERTRKTRTGEAFAAGLAPDCAITVHTQADTS 121

Qy 113 GDDPIFHGLDTG-C-LDOA--DAILAAGGGGLTARHG-----LTLAKVLNF--ASACL 159
Db 122 SPDFLFNPLKTVGCOLDNANVTDAILERAGSGTADFTGHYQTAFRELERVLNTPQSNLCL 181

Qy 160 -----ECARVGRGLPLLASTISEIFLLEFLLEYAQGPMEVHGRI 196
Db 182 KREKQDESCSLTQALPSELKVSADCVSLTGAVS--LASMLTEIFLLQQAQGMPEPGWRI 239

Qy 197 --SAEW---LLHNAQF-LANRTPYIA--RATPI--LIVTALS-----PARV 232
Db 240 TDSQWNTLLSLHNAQFDLQRTPEVARSRATPLDLIKTALTTPHPFQKQAYGVTLPTSV 299

Qy 233 VLLAGHDTNLA-LGG-LDL-WQLP-QPDETPPGGALVFELWNR-----YVRVMYQTM 280
Db 300 LFIAGHDTNLALGGALNLTLPQDPNTPPGGELVFERWRRLSDNSQWQVSLVFQTL 359

Qy 281 DQRLNLEPLPRILP-----IPGC---GSEACSLSDPARLV---APAC 318
Db 360 QQVRDRTPLSLNTPPGGVKLTLAGCERNAQGWCSLAGFTQIVNEARIPAC 410

RESULT 11
ABP51937

ID ABP51937 standard; protein; 410 AA.

XX ABP51937;
AC ABP51937;

XX 08-OCT-2002 (first entry)
XX Phytase protein sequence.

XX Phytase; enzyme; E.C.3.1.3.8; phytate; foodstuff; feed.
XX Unidentified.

XX WO200248332-A2.
XX 20-JUN-2002.

XX 12-DEC-2001; 2001WO-US048774.
XX 12-DEC-2000; 2000US-0255090P.

XX (DIVE-) DIVERSA CORP.
XX Short J, Mathur EJ, Richardson T, Robertson D, Barton N;

XX WPI; 2002-583504/62.
XX Novel recombinant phytase protein and polynucleotide for improving
PT nutritional value of phytate-containing foodstuff, in animal feed and
PT feed supplements and to degrade excess phytase from environment or
PT sample.

XX
PS
XX

Disclosure; Fig 9; 208pp; English.

CC The present invention describes an isolated phytase protein (I). (I) can
CC be used for improving the nutritional value of a phytate-containing
CC foodstuff, by contacting the phytate-containing foodstuff with (I), where
CC the enzyme catalyses the liberation of inorganic phosphate from the
CC phytate-containing foodstuff, and so improving the nutritive value of the
CC contacted foodstuff. The liberation of inorganic phosphate occurs prior
CC to or after the ingestion of phytate-containing foodstuff by a recipient
CC organism. Nucleotide sequences (II) encoding (I) can be used for
CC producing an animal feed, by transforming a plant, plant portion or plant
CC cell with a nucleic acid expression vector, comprising (II), culturing
CC the plant, plant portion or plant cell under conditions in which the
CC phytase protein is expressed and converting the plant, plant portion or
CC plant cell into a composition suitable for animal feed. The animal is
CC preferably a monogastric animal or a ruminant. (I) and (II) are useful in
CC animal feed and feed supplements as well as in treatments to degrade or
CC remove excess phytate from the environment or a sample. (I) reduces
CC phytate levels in animal manure and so reduces the phosphate pollution of
CC the environment. The present sequence represents a phytase from the
CC present invention

XX
SQ

Sequence 410 AA;

Query Match	37.2%	Score	613.5;	DB	5;	Length	410;
Best Local Similarity	45.5%;	Pred. No.	8.3e-52;				
Matches	186;	Conservative	41;	Mismatches	69;	Indels	113;
		Gaps	34;				
QY	21	AEP-LKLEKVVILSRHGVRSPTKA-LM-VSPD-WP-WPV---	GLTPRGAAVL-LIGY-68				
DB	2	SEPELKULESVVSRHGVRAPTKATQLMQDVTDPDAMPTWPKGLBELTPRGELIAYLGHY-61					
QY	69	---RYFARGLLP--GCPAAG--TILADVDERTR-TQQAFAAGLAPGCAI-VH----	AGD-114				
DB	62	WRQRLIVADGLLPKCGCPSQGVAIADVDERTKGTGEAFAAGLAPDCAITVHTQADTSSP-121					
QY	115	DPLEHGLDTG-C-LDQR---DAILAGEGLFARHG-----LTLAKLVNF--ASACL---	159				
DB	122	DPLFNPLKTGVCCQJDNANVTDAIDERAGGSIADFTGYQTAPRELVRVLPQSNLCKR-181					
QY	160	-----ECARVVGRGLPILLASTLSEIFLELYAAGPMPEVGMGRI--196					
DB	182	EKQDESCSLTQALPSELKVSADCVSLTGAVS--LASMLTEIFLLQQAQMPGPGWRITD-239					
QY	197	SAEW-----LLHNAQF-LMNTPTPIA-RATPI-LIVTALS-----PARVVL-234					
DB	240	SHOWNTLLSNAQFDLLQRTPEVARSATPFLDLIKTALTALTPHPQKOAYGVTLPTSVL-299					
QY	235	LAGHDTNLA-LGG-LDL-WOLP-QPDETTPGGALVPELWNR-----YRVVWQTMQD-282					
DB	300	IAGHDTNLANJGGALELNWTLPGQDNTPPGSELVEFRNRRLSDNSQWISQVSLVFQTLQ-359					
QY	283	LRNLEPLLRLP-----TPGC-----GSEAAACSLSDFARV-----APAC-318					
DB	360	MEDKTPISLNTFPPEVKLTLTLAGGEERNAQCMCSLAGFTQIVNEARIPAC-408					

RESULT 12	
AAB36257	
ID	AAB36257 standard; protein; 432 AA.
XX	
XX	
AC	AAB36257;
XX	
XX	
DT	12-SEP-2003 (revised)
DT	20-FEB-2001 (first entry)
XX	
XX	
DE	Lama2/APPA plasmid translated sequence.
XX	
XX	
KW	Transgenic animal; salivary protein; phytase; phosphorus; animal growth;
XX	environmental pollution; pig.
OS	
OS	Mus musculus.

OS	Escherichia coli.
OS	Chimeric.
XX	
PN	WO200064247-A1.
XX	
XX	02-NOV-2000.
XX	
XX	20-APR-2000; 2000WO-CA000430.
PF	
XX	
PR	23-APR-1999; 99US-0130508P.
XX	
PA	(UYGU-) UNIV GUELPH.
XX	
PI	Forsberg CW, Golovan S, Phillips JP;
XX	
DR	WPI; 2000-687245/67.
DR	N-PSDB; AAC68294.
XX	
PT	Transgenic non-human animal for gastrointestinal tract specific
PT	expression of a protein, preferably phytase, comprises a nucleic acid
PT	sequence including a heterologous transgene construct encoding the
PT	protein.
XX	
XX	Disclosure; Fig 5; 152pp; English.
XX	
CC	The present invention provides transgenic animals which produce desired
CC	proteins, in this case pigs which expresses phytase in the salivary
CC	gland. Low phytase production levels result in phytate in the diet being
CC	excreted and causing phosphorus contamination in water, as well as
CC	reducing the growth of animals. The invention provides a number of
CC	transgenes containing the E. coli APPA phytase coding sequence. (Updated
CC	on 12-SEP-2003 to standardise OS field)
XX	
SQ	Sequence 432 AA;

Query Match	37.2%;	Score	613.5;	DB	3;	Length	432;
Best Local Similarity	45.3%;	Pred. No.	8.9e-52;				
Matches	194;	Conservative	44;	Mismatches	79;	Indels	111;
						Gaps	36
QY	2	SLLAAALAAALAP	-AAAAAP-LKLEKWLIRSHGVSPTKA	-LM-	-VSPD-WP-W	52	
		:::	:::	:::	:::	:::	
Db	3	AILLPTLLPLTPQSAFAQSEPELKLESVIVSRHGVRAPTKATQLMODVTPDAMPWTW	62				
QY	53	PV---GLTPRGAAUV-LLGY----	RVFARGLL--PCCPAAG--TILADYDETR-TGQAF	99			
		:::	:::	:::	:::	:::	
Db	63	PVKLGWLTPRGELIAYLGHYQRLVADGLLAKKCPQSGGVAIADYDETRKTGEAF	122				
QY	100	AAGLAPGCAI-VH-----	AGDDPIFHGLDTG-C-LDQA---DAILAAGEGGILT--ARHGL	146			
		:::	:::	:::	:::	:::	
Db	123	RAGLAPDCAITVHTQATSPFDPLFPLKTVGVCOLDNANVTDAILSRAGGSIASTFTGHRQ	182				
QY	147	T-----LAKVLNF--ASACLE-----	CARVVGRIGPL-LASTLSEI	179			
		:::	:::	:::	:::	:::	
Db	183	TAFRELERVLNFPQSNCLCKREXQDBSCSLTQALPSELKVSADNVSLTGAVSLASMLTEI	242				
QY	180	FLLEVAQGEPMVGHGRIL--SAEW----	LLHNAQF-LMNETPVIA--RAPPI-L-LIWTALS	228			
		:::	:::	:::	:::	:::	
Db	243	FLLOQAQMPFEGHGRITDISHQWNTLLSLHNAQFYLLQRTPEVARSRAFLDLIKALT	302				
QY	229	-----PARVVLIAGHDTNIA-LGG-LDI-WOLP-QPDETPPGGALVPELMNR	271				
		:::	:::	:::	:::	:::	
Db	303	PHPPKQAYGVTLPTSFLVFIAGHDTNLNIGGALELNWTLPGQPDNTPPGGSELVFERWR	362				
QY	272	-----YVRVMYQTDQLRNLPEILLPRILP-----	IPGC---GSEAAACSLISDFARLV	314			
		:::	:::	:::	:::	:::	
Db	363	LSDNSQWLTQVSLVFQTLQWRDKTFLSLNTFPPEVKLTLAGEERNAQWMCSLAGETQIV	422				
QY	315	-----APAC	318				
		:::	:::				
Db	423	NEARIPAC	430				

RESULT 13

```
AAB36261
ID AAB36261 standard; protein; 432 AA.
XX
AC AAB36261;
XX
DT 12-SEP-2003 (revised)
DT 20-FEB-2001 (first entry)
XX
DE R15/APPA plasmid translated sequence.
XX
KW Transgenic animal; salivary protein; phytase; phosphorus; animal growth;
KW environmental pollution; pig.
XX
OS Rattus sp.
OS Escherichia coli.
OS Chimeric.
XX
PN WO200064247-A1.
XX
PD 02-NOV-2000.
XX
PF 20-APR-2000; 2000WO-CA000430.
XX
PR 23-APR-1999; 99US-0130508P.
XX
PA (UYGU-) UNIV GUELPH.
XX
PI Forsberg CW, Golovan S, Phillips JP;
XX
DR WPI; 2000-687245/67.
XX
DR N-PSDB; AAC68298.
XX
PT Transgenic non-human animal for gastrointestinal tract specific
PT expression of a protein, preferably phytase, comprises a nucleic acid
PT sequence including a heterologous transgene construct encoding the
PT protein.
XX
PS Disclosure; Fig 21; 152pp; English.
XX
CC The present invention provides transgenic animals which produce desired
CC proteins, in this case pigs which expresses phytase in the salivary
CC gland. Low phytase production levels result in phytate in the diet being
CC excreted and causing phosphorus contamination in water, as well as
CC reducing the growth of animals. The invention provides a number of
CC transgenes containing the E. coli APPA phytase coding sequence. (Updated
CC on 12-SEP-2003 to standardise OS field)
XX
SQ Sequence 432 AA;
Query Match 37.2%; Score 613.5; DB 3; Length 432;
Best Local Similarity 45.3%; Pred. No. 8.9e-52;
Matches 194; Conservative 44; Mismatches 79; Indels 111; Gaps 36;
QY 2 SLILALLALALAP--AAAAAP-LKLEKVVILSRHGVSRPTKA--LM--VSPD-WP-W 52
Db 3 ALLIPFLSLILPTPSAFQSEPELKLESVIVSRHGVRAPTKATQMDVTPDAPTW 62
QY 53 PV---GLTPRGAALV-LLGY---RYFARGLL--PGCPAAG--TILADVDERTR-TGQAF 99
Db 63 PVKLGWLTFRGGELIAYLGHYQORLVADGLLAKGCPQSGQVIAIADVDERTRKTGEAF 122
QY 100 AGLAPGCAI-VH-----AGDDPIFHGLDTG-C-LDQA---DAILAAGEGLT--ARHCL 146
Db 123 AGLAPDCAITVHTQADTSPPFLFPLKTVGVQLDQANVTAILSRAGGSADTFTGHRQ 182
QY 147 T-----LAKVINF--ASACLE-----CARVVGRGLPL-LASTLSEI 179
Db 183 TAPRELRVINFQSNCLKREKQDECSLTQALPSELKVSADNVSILTGAVALSMLTEI 242
QY 180 FLLEYAGPMVEVWGRI--SAEW---LLHNAQF-LMNETPYIA--RATPI--LIVTALS 228
Db 243 FLQQAQGMPEPGWGRITDSHQWNTLLSLHNAQFYLLQRTPEVARSATPFLDLIKTALT 302
```

```
229 -----PARVLLAGHDTNLA-IGG-LDL-WOLP-OPDETTPPGALVFELWNR 271
303 PHPPQKAYGVLTPTSVLPIAGHDTNLANLGGALELWNTLPQOPDNTTPGGELVFERWR 362
272 -----YVRVYQMDQLRNLEPLILPILP-----IPGC---GSEAACSLSDPARLV 314
363 LSDNSQWIVSLVFQTLQQRDKTPLSLNTPPGEVKLTLAGCERNAQGVCSLAGTQIV 422
315 ----APAC 318
423 NEARIPAC 430

RESULT 14
AAB36262
ID AAB36262 standard; protein; 432 AA.
XX
AC AAB36262;
XX
DT 12-SEP-2003 (revised)
DT 20-FEB-2001 (first entry)
XX
DE SV40/APPA plasmid translated sequence.
XX
KW Transgenic animal; salivary protein; phytase; phosphorus; animal growth;
KW environmental pollution; pig.
XX
OS Simian virus 40.
OS Escherichia coli.
OS Chimeric.
XX
PN WO200064247-A1.
XX
PD 02-NOV-2000.
XX
PF 20-APR-2000; 2000WO-CA000430.
XX
PR 23-APR-1999; 99US-0130508P.
XX
PA (UYGU-) UNIV GUELPH.
XX
PI Forsberg CW, Golovan S, Phillips JP;
XX
DR WPI; 2000-687245/67.
XX
DR N-PSDB; AAC68299.
XX
PT Transgenic non-human animal for gastrointestinal tract specific
PT expression of a protein, preferably phytase, comprises a nucleic acid
PT sequence including a heterologous transgene construct encoding the
PT protein.
XX
PS Disclosure; Fig 22; 152pp; English.
XX
CC The present invention provides transgenic animals which produce desired
CC proteins, in this case pigs which expresses phytase in the salivary
CC gland. Low phytase production levels result in phytate in the diet being
CC excreted and causing phosphorus contamination in water, as well as
CC reducing the growth of animals. The invention provides a number of
CC transgenes containing the E. coli APPA phytase coding sequence. (Updated
CC on 12-SEP-2003 to standardise OS field)
XX
SQ Sequence 432 AA;
Query Match 37.2%; Score 613.5; DB 3; Length 432;
Best Local Similarity 45.3%; Pred. No. 8.9e-52;
Matches 194; Conservative 44; Mismatches 79; Indels 111; Gaps 36;
QY 2 SLILALLALALAP--AAAAAP-LKLEKVVILSRHGVSRPTKA--LM--VSPD-WP-W 52
Db 3 ALLIPFLSLILPTPSAFQSEPELKLESVIVSRHGVRAPTKATQMDVTPDAPTW 62
QY 53 PV---GLTPRGAALV-LLGY---RYFARGLL--PGCPAAG--TILADVDERTR-TGQAF 99
```

```

Db 63 PVKGLMTPRGELIAYLGHYQRLVADGLLAKKGCPSQGVVAITADVDERTRKTGEAF 122
QY 100 AAGLAPCAI-VH-----AGDDPIFHGLDTG-C-LDOA-----DAILAAGEGLT--ARHGL 146
Db 123 AAGLAPCAITVHTQADTSFDPLEKLTGVCOLDNANVTDAILSRAGGSIAADFTGHRQ 182
QY 147 T-----LAKVNF--ASACLE-----CARVWGLGRLPL--LASTLSEI 179
Db 183 TAFRELERVNFPOSNLCRKQDESCSLTQALPSELKVSADNVSLTGAVSLASMLTEI 242
QY 180 FLLEYAOGPMVGVGRI--SAEW-----LHNAQF-LMNRTPYA--RATPI--LIVTALS 228
Db 243 FLQQOAGMPEPGWGRITDSHOWNTLLSLHNAQFYLLQRTPEVARSRAATPLDLIKTALT 302
QY 229 -----PARVLLAGHDTNLA-LGG-LDL-WOLP-OPDETTPPGALVPELWNR 271
Db 303 PHPPQKQAYGVTLFTSVLFIAGHDTNLANLGGALELNTLPGQDNTTPPGELVFERWR 362
QY 272 -----YVRVMYQTMQDLNLEPLPRILP-----IPGC-----GSEAACSLSDFARLV 314
Db 363 LSDNSQWISLVFQTLQMRDKTPLSLNTTPPGEVKLTLAGCEERNAQGMCSLAGFTQIV 422
QY 315 ----APAC 318
Db 423 NEARIPAC 430

RESULT 15
AAB36259
ID AAB36259 standard; protein; 432 AA.
AC AAB36259;
XX
DT 12-SEP-2003 (revised)
DT 20-FEB-2001 (first entry)
XX
DE
DE
DE
KW Transgenic animal; salivary protein; phytase; phosphorus; animal growth;
KW environmental pollution; pig.
XX
OS Rattus sp.
OS Escherichia coli.
OS Chimeric.
XX
PN WO2000064247-A1.
XX
XX
PD 02-NOV-2000.
XX
PF 20-APR-2000; 2000WO-CA000430.
XX
PR 23-APR-1999; 99US-0130508P.
XX
PA (UYGU-) UNIV GUELPH.
XX
PI Forsberg CW, Golovan S, Phillips JP;
XX
XX WPI; 2000-687245/67.
DR N-PSDB; AAC68296.
XX
XX Transgenic non-human animal for gastrointestinal tract specific
XX expression of a protein, preferably phytase, comprises a nucleic acid
XX sequence including a heterologous transgene construct encoding the
XX protein.
XX
XX Disclosure; Fig 19; 152pp; English.
XX
XX The present invention provides transgenic animals which produce desired
XX proteins, in this case pigs which expresses phytase in the salivary
XX gland. Low phytase production levels result in phytate in the diet being
XX excreted and causing phosphorus contamination in water, as well as
XX reducing the growth of animals. The invention provides a number of
XX transgenes containing the E. coli APPA phytase coding sequence. (Updated

```

```

CC on 12-SEP-2003 to standardise OS field)
XX
SQ Sequence 432 AA;
Query Match 37.2%; Score 613.5; DB 3; Length 432;
Best Local Similarity 45.3%; Pred. NO. 8.9e-52;
Matches 194; Conservative 44; Mismatches 79; Indels 111; Gaps 36;
QY 2 SLLLAALLAALAP-AAAAAEP-LKLEKVVILSRHGVRSPTKA--LM--VSPD-WP-W 52
Db 3 AILIFPLSLLPLTPQSFAQSEPELKLESVVIVSRHGVRAPTKATQLMQDVTDPDAPTW 62
QY 53 PV---GLTPRGALV-LLGY-----RYFARGLL--PGCPAAG--TILADVDERTR-TGQAF 99
Db 63 PVKGLMTPRGELIAYLGHYQRLVADGLLAKKGCPSQGVVAITADVDERTRKTGEAF 122
QY 100 AAGLAPCAI-VH-----AGDDPIFHGLDTG-C-LDOA-----DAILAAGEGLT--ARHGL 146
Db 123 AAGLAPCAITVHTQADTSFDPLEKLTGVCOLDNANVTDAILSRAGGSIAADFTGHRQ 182
QY 147 T-----LAKVNF--ASACLE-----CARVWGLGRLPL--LASTLSEI 179
Db 183 TAFRELERVNFPOSNLCRKQDESCSLTQALPSELKVSADNVSLTGAVSLASMLTEI 242
QY 180 FLLEYAOGPMVGVGRI--SAEW-----LHNAQF-LMNRTPYA--RATPI--LIVTALS 228
Db 243 FLQQOAGMPEPGWGRITDSHOWNTLLSLHNAQFYLLQRTPEVARSRAATPLDLIKTALT 302
QY 229 -----PARVLLAGHDTNLA-LGG-LDL-WOLP-OPDETTPPGALVPELWNR 271
Db 303 PHPPQKQAYGVTLFTSVLFIAGHDTNLANLGGALELNTLPGQDNTTPPGELVFERWR 362
QY 272 -----YVRVMYQTMQDLNLEPLPRILP-----IPGC-----GSEAACSLSDFARLV 314
Db 363 LSDNSQWISLVFQTLQMRDKTPLSLNTTPPGEVKLTLAGCEERNAQGMCSLAGFTQIV 422
QY 315 ----APAC 318
Db 423 NEARIPAC 430

Search completed: April 30, 2004, 12:38:48
Job time : 42.7195 secs

```

Sequence 3, Appli
Sequence 106, App
Sequence 128, App
Sequence 151, App
Sequence 12, Appl
Sequence 33, Appl
Sequence 33, Appl
Sequence 78, Appl
Sequence 6, Appli
Sequence 8, Appli
Sequence 9208, Ap
Sequence 28159, A
Sequence 18063, A
Sequence 27245, A
Sequence 28589, A
Sequence 19662, A
Sequence 19438, A
Sequence 17734, A

28 91.5 5.6 439 4 US-09-044-718-3
29 91.5 5.6 440 4 US-09-684-855-106
30 91.5 5.6 440 4 US-09-684-855-128
31 91.5 5.6 440 4 US-09-684-855-151
32 91.5 5.6 449 4 US-09-044-718-12
33 91.5 5.6 449 4 US-08-868-435-33
34 91.5 5.6 465 4 US-08-744-231-33
35 91.5 5.6 465 4 US-09-044-718-78
36 91.5 5.6 465 4 US-09-636-499-6
37 91.5 5.6 465 4 US-09-273-871A-8
38 91.5 5.6 499 4 US-09-489-039A-9208
39 91.5 5.6 560 4 US-09-252-931A-28159
40 91.5 5.6 579 4 US-09-252-931A-18063
41 91 5.5 580 4 US-09-252-931A-27245
42 91 5.5 841 4 US-09-252-931A-28589
43 91 5.5 1002 4 US-09-252-931A-19662
44 90.5 5.5 461 4 US-09-252-931A-19438
45 90 5.5 581 4 US-09-252-931A-17734

ALIGNMENTS

RESULT 1

US-09-259-214-2
; Sequence 2, Application US/09259214A
; Patent No. 6110719
; GENERAL INFORMATION:
; APPLICANT: Kretz, Keith
; TITLE OF INVENTION: NOVEL PHYTASE
; FILE REFERENCE: DIVER1370-1
; CURRENT APPLICATION NUMBER: US/09/259,214A
; CURRENT FILING DATE: 1999-03-01
; EARLIER APPLICATION NUMBER: 08/910,798
; EARLIER FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-259-214-2

Query Match 37.1%; Score 611.5; DB 3; Length 440;
Best Local Similarity 45.1%; Pred. No. 1.5e-59;
Matches 193; Conservative 45; Mismatches 79; Indels 111; Gaps 36;
QY 2 SLLLAALLAALAP--AAAAEP-1KLEKVVILSRHGVSRPTKA--LM--VSPD-WP-W 52
DB 3 ALIPLSLILPLTPQSAFAQSEPELKLESVVIVSRHGVRAPTKATQTLQMDVTPDAWPTW 62
QY 53 PV---GLTPRGARLV-LLGY-----RYFARGLL--PCPPAAG--TILADVDETR-TGQAF 99
DB 63 PVKLGWLTFRGGELIAYLGHYQRLVADGLAKKGCPOSGQVAILADVDETRKTGEAF 122
QY 100 AAGLAPGCAI-VH-----AGDDPIPHGLDTG-C-LDQA---DAILAAGSGGLT--ARHGL 146
DB 123 AAGLAPDCAITVHTQADTSSPDPLFNPLKTVGCQLDNANVTAILSRAGGSIAFTGHQ 182
QY 147 T-----LAKVINF--ASACLE-----CARVVGRLGPL-LASTLSI 179
DB 183 TARELERVLNFPQSNLCRKQDSSCSLTQALPSELKVSADNVSLTGAVSLASMLTEI 242
QY 180 FLLEYAQGPMEVWGRI--SAEW-----ILHNAQF-IMNTPYIA--RATPI--LIVTALS 228
DB 243 FLIQAQGMPEPGWGRI--SHQNTLSLHNAQFVLLQRTPEVARSRATPLDLIWAALT 302
QY 229 -----PARVVLIGHDTNLA-IGG-LDL-WQLP-QPDETTPFGALVFLWNR 271
DB 303 PHFPQKAYGVLTPTSVLFIAGHTNTNLNGLGALNLTLPQPDNTPFGGLVFFWR 362
QY 272 -----YRVVYQNDQLRNLPLPLP-----IPGC---GSEACSLSDFARLV 314

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 30, 2004, 12:36:16 ; Search time 12.7422 Seconds
(without alignments)
1288.405 Million cell updates/sec

Title: US-10-021-723B-14

Perfect score: 1647

Sequence: 1 VSLLAALLAALAPAAA.....CGSEACSLSDFARLVAPAC 318

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*

2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*

3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*

4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*

5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*

6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	611.5	37.1	440	3	US-09-259-214-2
2	611.5	37.1	440	3	US-09-318-528-2
3	611.5	37.1	440	3	US-09-291-931-2
4	608.5	36.9	433	4	US-09-540-149A-1
5	556	33.8	423	2	US-08-910-798-2
6	272.5	16.5	421	4	US-09-489-039A-7512
7	265.5	16.1	522	4	US-09-489-039A-13501
8	116	7.0	798	4	US-09-252-931A-25309
9	103.5	6.3	386	1	US-08-758-213-1
10	103.5	6.3	386	2	US-08-692-787-48
11	103.5	6.3	386	3	US-09-097-199-48
12	103.5	6.3	514	4	US-09-252-931A-25845
13	103.5	6.3	515	2	US-09-146-283-2
14	103.5	6.3	515	3	US-08-579-823A-2
15	103.5	6.3	515	3	US-09-324-185-2
16	102.5	6.2	4551	3	US-09-340-878-1
17	102.5	6.2	4551	4	US-09-141-908-2
18	102.5	6.2	4551	4	US-09-657-440-1
19	102.5	6.2	4613	3	US-09-105-537-31
20	102.5	6.2	11877	3	US-09-105-537-6
21	99	6.0	638	4	US-09-252-931A-18036
22	96.5	5.9	465	4	US-09-044-718-79
23	93.5	5.7	318	3	US-08-680-506-3
24	93.5	5.7	449	3	US-08-680-506-7
25	93.5	5.7	831	4	US-09-252-931A-30097
26	92.5	5.6	281	4	US-09-252-931A-19858
27	92.5	5.6	467	4	US-09-684-855-165

Db 363 LSDNSQWVQVSLVFTLQMRDKTPLSLNTPPGEVKLTAGCERNAGMCSLAGFTQIV 422
 QY 315 ----APAC 318
 Db 423 NEARIPAC 430

RESULT 2

US-09-318-528-2
 ; Sequence 2, Application US/09318528
 ; Patent No. 6183740
 ; GENERAL INFORMATION:
 ; APPLICANT: Kretz, Keith
 ; TITLE OF INVENTION: NOVEL PHYTASE
 ; FILE REFERENCE: 09010/029003
 ; CURRENT APPLICATION NUMBER: US/09/318,528
 ; CURRENT FILING DATE: 1999-05-25
 ; EARLIER APPLICATION NUMBER: 09/291,931
 ; EARLIER FILING DATE: 1999-04-13
 ; EARLIER APPLICATION NUMBER: 08/910,798
 ; EARLIER FILING DATE: 1997-08-13
 ; EARLIER APPLICATION NUMBER: 09/259,214
 ; EARLIER FILING DATE: 1999-03-01
 ; NUMBER OF SEQ ID NOS: 4
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 2
 ; LENGTH: 440
 ; TYPE: PRT
 ; ORGANISM: Escherichia coli
 ; US-09-318-528-2

Query Match 37.1%; Score 611.5; DB 3; Length 440;
 Best Local Similarity 45.1%; Pred. No. 1.5e-59;
 Matches 193; Conservative 45; Mismatches 79; Indels 111; Gaps 36;

QY 2 SLLAALLAALAP--AAAAEP--IKLEKVVILSRHGVSPYKA--LM--VSPD-WP-W 52
 Db 3 AILIPFLSLIPLTPQSAFAQSBPELKESVIVSRHGVRAPTKATQMLQMDVTPDAPWTW 62
 QY 53 PV---GLTPRGAALV--LLGY---RYFARGLL--PGCPAAG--TILADVDERT--TQCAF 99
 Db 63 PVKLGWLTTPRGGLIAYLGHVQRLVADGLLAKKGCPOSGQVAILADVDERTKTGEAF 122
 QY 100 AAGLAPGCAI-VH-----AGDDPIFHGLDTG-C-LDQA---DALLAGEGGLT--ARHGL 146
 Db 123 AAGLAPDCAITVHTQADTSSPDPLFNPLKTGVQOLDNANVTDAILSRAGGSIAFTGHRQ 182
 QY 147 T---LAKVLF--ASACLE-----CARVVGRLGPL--LASTLSEI 179
 Db 183 TAFRELERVLNFPQSNCLKREKQDESCSLTQALPSELKVSADNVSLTGAVSLASMLTEI 242
 QY 180 FLLEYAQGPMEVGVGRI--SAEW-----LLHNAQF-LMNRTPYTA--PATPI--LIVTALS 228
 Db 243 FLQQAQGMPEPGWGRITDHSQWNTLLSLHNAQFYLLQRTPEVARSEATPLDLIMAALT 302
 QY 229 -----PARVLLAGHDNTLA-LGG-LDL-WOLP-OPDETTPGGALVPELWNR 271
 Db 303 PHPPQKAYGVTLPTSVLFTAGHDTNLNAGALELNLWTLPGQPDNTPPGGELVFERWR 362
 QY 272 -----YRVVMYQTMQDLRNLEPLLRILP-----IPGC---GSEAAACSLSDPARLV 314
 Db 363 LSDNSQWVQVSLVFTLQMRDKTPLSLNTPPGEVKLTAGCERNAGMCSLAGFTQIV 422
 QY 315 ----APAC 318
 Db 423 NEARIPAC 430

RESULT 3

US-09-291-931-2
 ; Sequence 2, Application US/09291931A
 ; Patent No. 6190897
 ; GENERAL INFORMATION:

; APPLICANT: Kretz, Keith
 ; TITLE OF INVENTION: NOVEL PHYTASE
 ; FILE REFERENCE: 09010/029003
 ; CURRENT APPLICATION NUMBER: US/09/291,931A
 ; CURRENT FILING DATE: 1999-04-13
 ; EARLIER APPLICATION NUMBER: 08/910,798
 ; EARLIER FILING DATE: 1997-08-13
 ; EARLIER APPLICATION NUMBER: 09/259,214
 ; EARLIER FILING DATE: 1999-03-01
 ; NUMBER OF SEQ ID NOS: 4
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 2
 ; LENGTH: 440
 ; TYPE: PRT
 ; ORGANISM: Escherichia coli
 ; US-09-291-931-2

Query Match 37.1%; Score 611.5; DB 3; Length 440;
 Best Local Similarity 45.1%; Pred. No. 1.5e-59;
 Matches 193; Conservative 45; Mismatches 79; Indels 111; Gaps 36;

QY 2 SLLAALLAALAP--AAAAEP--IKLEKVVILSRHGVSPYKA--LM--VSPD-WP-W 52
 Db 3 AILIPFLSLIPLTPQSAFAQSBPELKESVIVSRHGVRAPTKATQMLQMDVTPDAPWTW 62
 QY 53 PV---GLTPRGAALV--LLGY---RYFARGLL--PGCPAAG--TILADVDERT--TQCAF 99
 Db 63 PVKLGWLTTPRGGLIAYLGHVQRLVADGLLAKKGCPOSGQVAILADVDERTKTGEAF 122
 QY 100 AAGLAPGCAI-VH-----AGDDPIFHGLDTG-C-LDQA---DALLAGEGGLT--ARHGL 146
 Db 123 AAGLAPDCAITVHTQADTSSPDPLFNPLKTGVQOLDNANVTDAILSRAGGSIAFTGHRQ 182
 QY 147 T---LAKVLF--ASACLE-----CARVVGRLGPL--LASTLSEI 179
 Db 183 TAFRELERVLNFPQSNCLKREKQDESCSLTQALPSELKVSADNVSLTGAVSLASMLTEI 242
 QY 180 FLLEYAQGPMEVGVGRI--SAEW-----LLHNAQF-LMNRTPYTA--PATPI--LIVTALS 228
 Db 243 FLQQAQGMPEPGWGRITDHSQWNTLLSLHNAQFYLLQRTPEVARSEATPLDLIMAALT 302
 QY 229 -----PARVLLAGHDNTLA-LGG-LDL-WOLP-OPDETTPGGALVPELWNR 271
 Db 303 PHPPQKAYGVTLPTSVLFTAGHDTNLNAGALELNLWTLPGQPDNTPPGGELVFERWR 362
 QY 272 -----YRVVMYQTMQDLRNLEPLLRILP-----IPGC---GSEAAACSLSDPARLV 314
 Db 363 LSDNSQWVQVSLVFTLQMRDKTPLSLNTPPGEVKLTAGCERNAGMCSLAGFTQIV 422
 QY 315 ----APAC 318
 Db 423 NEARIPAC 430

RESULT 4

US-09-540-149A-1
 ; Sequence 1, Application US/09540149A
 ; Patent No. 6511699
 ; GENERAL INFORMATION:
 ; APPLICANT: Lei, Xingen
 ; TITLE OF INVENTION: ENZYMES WITH IMPROVED PHYTASE ACTIVITY
 ; FILE REFERENCE: 19603/2791
 ; CURRENT APPLICATION NUMBER: US/09/540,149A
 ; CURRENT FILING DATE: 2000-03-31
 ; PRIOR APPLICATION NUMBER: 60/127,032
 ; PRIOR FILING DATE: 1999-03-31
 ; NUMBER OF SEQ ID NOS: 9
 ; SOFTWARE: Patent in Ver. 2.1
 ; SEQ ID NO 1
 ; LENGTH: 433
 ; TYPE: PRT
 ; ORGANISM: Escherichia coli
 ; FEATURE:

NAME/KEY: UNSURE
LOCATION: (433)
OTHER INFORMATION: Xaa at position 433 in this sequence is unknown
US-09-540-149A-1

Query Match 36.9%; Score 608.5; DB 4; Length 433;
Best Local Similarity 45.1%; Pred. No. 3.1e-59;
Matches 193; Conservative 43; Mismatches 81; Indels 111; Gaps 36;

QY 2 SLLAALLAALAP--AAAAEP-LKLEKVVILSRHGVRSPTKA--LM--VSPD-WP-W 52
DB 3 AILIPFSLIPLTPQSAFAQSEPELKLESVIVSRHGVRAPTKATQIMQDVTDPANPTW 62
QY 53 PV---GLTPGGAALV-LLGY---RYFARGLL--PGCPAAG--TILADYDERTR-TGOAF 99
DB 63 PVKLGWLTFRGGELIAYLGHYQORLVADGILLAKKCPQPGQVAIADVDERTKTKQAD 122
QY 100 AAGLAPGCAI-VH-----AGDDPIFRGLDTG-C-LDOA---DAILAAGEGGLT--ARHGL 146
DB 123 AAGLAPGCAITVHTQADTSSPDLFNLKTVGCOLDNANVTDAILSRAGGSIAADFTGHRQ 182
QY 147 T-----LAKVLNEA--SACLE-----CARVVGRLGFL--LASTLSEI 179
DB 183 TAFRELVLNFSQNLNLRKQDESCSLTQALPSELKVSADNLSLTGAVSLASMLTEI 242
QY 180 FLLEYAQGPMEVGVGRI--SAEW-----LLHNAOF-LMNRTPYIA--RATPI--LIVTALS 228
DB 243 FLLOQAQGMPEPCWGRIITDQSHQNTLLSLHNAQFYLLQRTPEVARSRATPLDLIQAALT 302
QY 229 -----PARVLLAGHDNTLA-LGG-LDL-WQLP-OPDETTPGGALVFLWNR 271
DB 303 PHPPQKQAVGVTUFTSVLFTAGHDNTLANLGGALELNLWTLPGQPDNTTPGGELVFERWR 362
QY 272 -----YRVVYQTMQDLNLEPLLRILP-----IPGC---GSEAAACSLSDFARLV 314
DB 363 LSDNSQWIVSLVFTLQQRDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGTQIV 422
QY 315 -----APAC 318
DB 423 NEARIPAC 430

RESULT 5

US-08-910-798-2
Sequence 2, Application US/08910798
Patent No. 5876997
GENERAL INFORMATION:
APPLICANT: KRETTZ
TITLE OF INVENTION: NOVEL PHYTASE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: California
COUNTRY: US
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/910,798
FILING DATE: August 13, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: HAILE, PH.D., Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/029001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 423 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-910-798-2

Query Match 33.8%; Score 556; DB 2; Length 423;

Best Local Similarity 43.1%; Pred. No. 2.2e-53;
Matches 182; Conservative 44; Mismatches 80; Indels 116; Gaps 35;

QY 2 SLLAALLAALAP--AAAAEP-LKLEKVVILSRHGVRSPTKA--LM--VSPD-WP-W 52
DB 3 AILIPFSLIPLTPQSAFAQSEPELKLESVIVSRHGVRAPTKATQIMQDVTDPANPTW 62
QY 53 PV---GLTPGGAALV-LLGY---RYFARGLL--PGCPAAG--TILADYDERTR-TGOAF 99
DB 63 PVKLGWLTFRGGELIAYLGHYQORLVADGILLAKKCPQSGQVAIADVDERTKTKQAD 122
QY 100 AAGLAPGCAIVHAGDDPIFRGLDTG-C-LDOA---DAILAAGEGGLT--ARHGLT---L 148
DB 123 T-----SSPDLFNLKTVGCOLDNANVTDAILSRAGGSIAADFTGHRQTAFREL 171
QY 149 AKVLNF--SACLE-----CARVVGRLGFL--LASTLSEIFLLEYA 185
DB 172 ERLNFPQSNLCKRKQDESCSLTQALPSELKVSADNLSLTGAVSLASMLTEIFLLQQA 231
QY 186 QGPMEVGVGRI--SAEW-----LLHNAOF-LMNRTPYIA--RATPI--LIVTALS--- 228
DB 232 QGMPEPGMGRITDQSHQNTLLSLHNAQFYLLQRTPEVARSRATPLDLIQAALTTPHPPOK 291
QY 229 -----PARVLLAGHDNTLA-LGG-LDL-WQLP-OPDETTPGGALVFLWNR----- 271
DB 292 QAYGVTUFTSVLFTAGHDNTLANLGGALELNLWTLPGQPDNTTPGGELVFERWRRLSDNSQ 351
QY 272 --YRVVYQTMQDLNLEPLLRILP-----IPGC---GSEAAACSLSDFARLV----AP 316
DB 352 WIQVSLVFTLQQRDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGTQIVNEARIP 411
QY 317 AC 318
DB 412 AC 413

RESULT 6

US-09-489-039A-7512
Sequence 7512, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 2709.2004001
CURRENT APPLICATION NUMBER: US/09/489,039A
CURRENT FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 7512
LENGTH: 421
TYPE: PRT
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-7512

Query Match 16.5%; Score 272.5; DB 4; Length 421;

Best Local Similarity 28.0%; Pred. No. 9.6e-22;
Matches 114; Conservative 62; Mismatches 130; Indels 101; Gaps 26;

QY 6 AAILAALAPAAAAEP-LKLEKVVILSRHGVRSPTKA--LM--VSPD-WP-WPVG- 55
DB 19 AVMLSAGAAQDKAAPGEGYQLQVILMSRNLRLAPLANSVLEQSTAKAWPQWDVFGGQ 78

```

QY 56 LTPRGAAL-VLLGY---RYFARGLL---PGCPAAGTILADVDERTR---TQQAFAAGLAP 105
Db 79 LTTKGVLEVNGHVMRWLAQKQLVTSGECPPENAVYANSQRTVATAQFPIITGAP 138
QY 106 GCAIVHAGDDPIFHGLDTGCLDQ-----ADAILAAGEGLTA-----RHGLTLAK----- 150
Db 139 GCGI-----PVHQPOGMTDPTFNPVITDDSPAFREKALQAMKEKQGMOLTESYKLL 192
QY 151 --VLNF--ASACLE--CARVVGR-----LGPL--LASTISEIFLLEYAOG-- 187
Db 193 ETMIDYRNSPSCKEKKVCSLEGKDTFSAGYQQPQVSGPLKVGNSLVDAFTLQYVSGFP 252
QY 188 PNEVWGGRISA--EW-----LHNAQFLMNTPTIAR--ATPI-----LIVTALSAPAR 231
Db 253 KDQVAMGEIADSKQWVLSKLNGYQDSLFTSVAVQNVAKPLVKYIDNALVGECAKAK 312
QY 232 VVLLAGHTNLA--LGGDL--WOLP--QPDSTPPGGLVFFELW-----NR---YRVVYQ 278
Db 313 VVLLVGHUSNIASLTALDFFPYQLPGQYERTPIGKLLFRWHDSSAGNRDLMKIEVYQ 372
QY 279 TMDQLENTLEPLPRILP-----IPGC--GSEACSLSDFAFLVAPA 317
Db 373 STEQLRNADALTQAPPQVRVTLALNGCPVDQDQFCPLETFKRVINEA 419

```

RESULT 7

US-09-489-039A-13501
; Sequence 13501, Application US/09489039A
; Patent No. 6610836

GENERAL INFORMATION:
APPLICANT: Gary Breton et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 2709.2004001

CURRENT APPLICATION NUMBER: US/09/489,039A

CURRENT FILING DATE: 2000-01-27

PRIOR APPLICATION NUMBER: US 60/117,747

PRIOR FILING DATE: 1999-01-29

NUMBER OF SEQ ID NOS: 14342

SEQ ID NO 13501

LENGTH: 522

TYPE: PRT

ORGANISM: Klebsiella pneumoniae

US-09-489-039A-13501

Query Match 16.1%; Score 265.5; DB 4; Length 522;

Best Local Similarity 30.5%; Pred. No. 8e-21; Indels 85; Gaps 24;
Matches 119; Conservative 44; Mismatches 142

```

QY 3 LLLAALLAALAPAAAAEPLKLEKWLTSRHGVRSP-----KALMVSPDWPV----- 54
Db 112 LFTACALPLLALQSA---DWQLEKVVLSRHGIRPPTAGNREAREATGSPTEWTT 169
QY 55 --GLTPRGALVL-----LGYYFARGLL--PGCPAAGTIL--ADVDERTR--TGOAFAAGL 103
Db 170 DGLTGHGAAYVNGKRAEQHVTQGLGQAGCTAHSIYVRASPLQRTAQAALVDA 229
QY 104 APOCAI--VH--AGD--DPIFHGLDTGCLDQAD-----AIIAAGEGGTARHGJTAKVLNF 154
Db 230 FPGGVAIHVSGDADPLFQ--TDKFAATQTDPAQLAAVKEKAGDLAQRQALAPTQL 288
QY 155 ASACLECARV-----GRLGPLLASTLS-----EIFLLEYAOG--PW--EYWGMR 195
Db 289 KQAVCOADKPCPIFDTPWQVEQSKGKTITISGVNANVETLRLGWSNMLPSQLAWK 348
QY 196 ISAEW-----LHNAQFLMNTPTIARATPIIVTAL-----SP--ARVVLLAGHD 239
Db 349 IQQAQITALLPLLTENDVLTNDVLYTAQKSGVLLNAMLGDGVKPEANPNRWLLVAHD 408
QY 240 TNALGGLDL---WOLP--QPDSTPPGGLVFFELW-----NRVYVYQ--MDQLRN 286
Db 409 TNIAMVATLNFWSQLPGYGRGNIPPGSSLVLEWRNNAKSGERYLRYFOAQGLDLRL 468

```

```

QY 287 E-----PULPRILPIPGCGSEACSLSDP 310
Db 469 QTPDAQHMLRQEWHPQCGCRQDVTGTLCPF 498

```

RESULT 8

US-09-252-991A-25309

; Sequence 25309, Application US/09252991A

; Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 25309

LENGTH: 798

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-25309

Query Match 7.0%; Score 116; DB 4; Length 798;

Best Local Similarity 27.0%; Pred. No. 0.00074; Indels 66; Gaps 12;

Matches 77; Conservative 24; Mismatches 118

```

QY 67 GY-----RVFARGLLPGCPAAGTILADVDERTRTG-----QAFAGLAPGCAIVHA 112
Db 486 GYDFVRRRTWASDVLEVPGGRLAALPELIEPGACIGNLRPEAAALGLAPHTRVACG 545
QY 113 GDDPIFHGLDTG-----CLDQADAILAAGEGGLTARHG-----LTLAKV 151
Db 546 GGDNMLAAIGTGNIRPCLLTASIGTSLSAVAERPLVPHGLATFCASSGGWLPACT 605
QY 152 LNFASACLECARVVGRIGLPLLASTLSEIFLLEYAOGFMEVGMGRISAEWLLHNAQFLMNR 211
Db 606 MNLTGAC-----GLVQDLHLDLDFEFLA--AQAP-----VGAEGLLMLPFDGER 650
QY 212 TPVIARATPILI-VTALSAPRVLLAGHTNLTALG---GLDLWQ---LQPDSTPPGAL 264
Db 651 VPALPHASASLHGWTANLSRANLCRAVLGTAFLGRYGLDLIRASGLPGEIRLVGGAA 710
QY 265 VFELMNVYRVVYQTMQDLRNLEPLPRILPIPGCGS--EAACSL 307
Db 711 KNPLWR-----TLADILGLPLVCPQTEAAALGAALQAANSL 748

```

RESULT 9

US-08-758-213-1

; Sequence 1, Application US/08758213

; Patent No. 5763490

GENERAL INFORMATION:

APPLICANT: Lebioda, L.

APPLICANT: Jakob, C.

TITLE OF INVENTION: TREATING PROSTATE CANCER BY INHIBITING PROSTATIC ACID PHOSP

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: Michael A. Mann, P.A.

STREET: Post Office Box 7908

CITY: Columbia

STATE: South Carolina

COUNTRY: United States

ZIP: 29202

COMPUTER READABLE FORM: On attached diskette

MEDIUM TYPE: Diskette 3.5 inch, 1.4 MB storage

COMPUTER: Apple Macintosh TM

OPERATING SYSTEM: Macintosh TM

SOFTWARE: Microsoft Word 5.0 TM

CITY: Houston
STATE: Texas
COUNTRY: United States of America
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 20-MAY-1996
APPLICATION NUMBER: 08/309,091
FILING DATE: 09/20/94
ATTORNEY/AGENT INFORMATION:
NAME: Mann, Michael A.
REGISTRATION NUMBER: 32,825
REFERENCE/DOCKET NUMBER: 96-2578
TELECOMMUNICATION INFORMATION:
TELEPHONE: (803) 254-8472
TELEFAX: (803) 254-3760
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 386 amino acids
TYPE: Amino acid
STRANDEDNESS: No. 5763490 Applicable
TOPOLOGY: Linear
US-08-758-213-1

Query Match 6.3%; Score 103.5; DB 1; Length 386;
Best Local Similarity 23.8%; Pred. No. 0.0062;
Matches 92; Conservative 38; Mismatches 148; Indels 109; Gaps 20;

QY 2 SLIIAALLAALAPAAAAAEPLKLVILSRHGVSPPTKALMVSP--DWPWPVG---L 56
Db 13 SLSLGLFLFFWLDRLSVLAKELF--VTLVFRHGDRSPIDTPTDPIKSSWPQGFQGL 70
QY 57 TPRG-----AALVILGYRYF-----ARGLLPGCPAAGTILADVDRTRTGOA 98
Db 71 TQLGMEQHYELGEYIRKRYKFLNESYKHEQVYIRS-----TDVDRTLMSAMT 118
QY 99 FAAGLAP--GCAI-----VHA-----GDD-----PIFHGLDTGCLDQ--- 128
Db 119 NLALFPPEGVSIWNPILLMQPIFVHTVPLSEDLQLLYPFRNCPRFQLESETLKSEEFQ 178
QY 129 -----ADAILAAGEGGLTARHGLTLAKVLNFAACLECARVVGRLGPIILAS----- 174
Db 179 KRLHPYKDFIATLKG--LSGLHGQDLFGIWSKVYDPLCYESVHNFTLPSWATEDWTMKLR 236
QY 175 TLSIFILLEYAOGPMEVWGGRISAEWLLHNAQFLMNRTPYIARATPILIVTALSAPARVL 234
Db 237 ELSLSLLSLYGIHKQEKSLQGGVLNE---ILN---HMKRATQI-----PSYKKLIM 285
QY 235 LAGHDTNLALGGLDLWQLPQDPETPPGGA-----LVFELMNRVYRVVYQTMQDLRNLEPL 289
Db 286 YSAHDTTVS--GLQWALDVYNGLLPPYASCHLTLYFEKGFEYFVEMY-----RNETQH 337
QY 290 LPRILPIPGCGSEACSLSDPARLVAP 316
Db 338 EYPVLMPLGC--SPSCPLERFAELVGP 362

RESULT 10
US-08-692-787-48

; Sequence 48, Application US/08692787
; Patent No. 582864
; GENERAL INFORMATION:
; APPLICANT: Ar, Gang
; APPLICANT: O'Hara, S. Mark
; APPLICANT: Ralph, David
; APPLICANT: Velti, Robert
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
; TITLE OF INVENTION: PROGNOSIS AND MANAGEMENT OF PROSTATE
; TITLE OF INVENTION: DISEASE
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433

CITY: Houston
STATE: Texas
COUNTRY: United States of America
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 09/20/94
APPLICATION NUMBER: US/08/692,787
FILING DATE: Concurrently Herewith
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Corider, Timothy S.
REGISTRATION NUMBER: 38,414
REFERENCE/DOCKET NUMBER: UROC:012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 386 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLSCULE TYPE: Protein
US-08-692-787-48

Query Match 6.3%; Score 103.5; DB 2; Length 386;
Best Local Similarity 23.8%; Pred. No. 0.0062;
Matches 92; Conservative 38; Mismatches 148; Indels 109; Gaps 20;

QY 2 SLIIAALLAALAPAAAAAEPLKLVILSRHGVSPPTKALMVSP--DWPWPVG---L 56
Db 13 SLSLGLFLFFWLDRLSVLAKELF--VTLVFRHGDRSPIDTPTDPIKSSWPQGFQGL 70
QY 57 TPRG-----AALVILGYRYF-----ARGLLPGCPAAGTILADVDRTRTGOA 98
Db 71 TQLGMEQHYELGEYIRKRYKFLNESYKHEQVYIRS-----TDVDRTLMSAMT 118
QY 99 FAAGLAP--GCAI-----VHA-----GDD-----PIFHGLDTGCLDQ--- 128
Db 119 NLALFPPEGVSIWNPILLMQPIFVHTVPLSEDLQLLYPFRNCPRFQLESETLKSEEFQ 178
QY 129 -----ADAILAAGEGGLTARHGLTLAKVLNFAACLECARVVGRLGPIILAS----- 174
Db 179 KRLHPYKDFIATLKG--LSGLHGQDLFGIWSKVYDPLCYESVHNFTLPSWATEDWTMKLR 236
QY 175 TLSIFILLEYAOGPMEVWGGRISAEWLLHNAQFLMNRTPYIARATPILIVTALSAPARVL 234
Db 237 ELSLSLLSLYGIHKQEKSLQGGVLNE---ILN---HMKRATQI-----PSYKKLIM 285
QY 235 LAGHDTNLALGGLDLWQLPQDPETPPGGA-----LVFELMNRVYRVVYQTMQDLRNLEPL 289
Db 286 YSAHDTTVS--GLQWALDVYNGLLPPYASCHLTLYFEKGFEYFVEMY-----RNETQH 337
QY 290 LPRILPIPGCGSEACSLSDPARLVAP 316
Db 338 EYPVLMPLGC--SPSCPLERFAELVGP 362

RESULT 11

US-09-097-199-48
; Sequence 48, Application US/09097199
; Patent No. 6218529
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; APPLICANT: O'Hara, S. Mark
; APPLICANT: Ralph, David
; APPLICANT: Velti, Robert
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
; TITLE OF INVENTION: PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
; NUMBER OF SEQUENCES: 87

;; CORRESPONDENCE ADDRESS:
;; ADDRESSES: Arnold, White & Durkee
;; STREET: P.O. Box 4433
;; CITY: Houston
;; STATE: Texas
;; COUNTRY: USA
;; ZIP: 77210
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent in Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/097,199
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/692,787
;; FILING DATE: 31-JUL-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Nakaehima, Richard A.
;; REGISTRATION NUMBER: P-42,023
;; REFERENCE/DOCKET NUMBER: UROC:018
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (512) 418-3000
;; TELEFAX: (512) 474-7577
;; INFORMATION FOR SEQ ID NO: 48:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 386 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-09-097-199-48

Query Match 6.3%; Score 103.5; DB 3; Length 386;
Best Local Similarity 23.8%; Pred. No. 0.0062;
Matches 92; Conservative 38; Mismatches 148; Indels 109; Gaps 20;
QY 2 SLIAAALLAALAPAAAAEPLKLVILSRHGVSRPTKALMVSP--DWPVPV--L 56
Db 13 SLISGLFLFLFWLDRSLAKELKF--VTLNFRHGDRSPIDFTDPKESWPQGGQL 70
QY 57 TPRG-----AALVLGVYRP-----AGLLPGCPAAGTILADVTRTGOA 98
Db 71 TQGMQHVGLGYIKRYKRYKFLNESYKHEQVIRS-----TDVTRTMSAMT 118
QY 99 FAAGLAP--GCAT-----VHA--GDD-----PIFHGDTGCLDQ---- 128
Db 119 NLAAALPPGVSINWPIILLWQPIVHTVPLSEDDQLLYLPPRCPRFOESETLKSEEPQ 178
QY 129 -----ADAILAAGEGGLTARHGLTLAKVLPASACLECARVVGRLGPLAS----- 174
Db 179 KRLHPYKDFATLIGK--LSGLHQDQLFGIWSKYVDLYCESVHNFTLPSWATEDTWTUR 236
QY 175 TLSEIFLLEVAQGPMEYGVGRISAEMLLHNAQFLMNRTPYIARATPILIVTALSFARVVL 234
Db 237 ELSELSSLSLYGIHKQKESRLOGGVLVNE--ILN--HMKEATQI-----PSYKLLIM 285
QY 235 LAGHDNTNALGGDLMLQDPDPTPPGGA-----LVFELNRYRVYVYQTMOLRLNLEPL 289
Db 286 YSARDITVS--GLQMALDVYNGLLPPYASCHLTLYFEKGEYFVEMY-----RNETHQ 337
QY 290 LPRILPPIPGGSEAAACSLSDFAILVAP 316
Db 338 EPIFLMLPGC--SPSCFLERFAELVGP 362

RESULT 12
US-09-252-991A-25845
; Sequence 25845, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.

;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;; FILE REFERENCE: 107196.136
;; CURRENT APPLICATION NUMBER: US/09/252,991A
;; CURRENT FILING DATE: 1999-02-18
;; PRIOR APPLICATION NUMBER: US 60/074,788
;; PRIOR FILING DATE: 1998-02-18
;; PRIOR APPLICATION NUMBER: US 60/094,190
;; PRIOR FILING DATE: 1998-07-27
;; NUMBER OF SEQ ID NOS: 33142
;; SEQ ID NO 25845
;; LENGTH: 514
;; TYPE: PRT
;; ORGANISM: Pseudomonas aeruginosa
;; US-09-252-991A-25845

Query Match 6.3%; Score 103.5; DB 4; Length 514;
Best Local Similarity 24.7%; Pred. No. 0.0095;
Matches 96; Conservative 37; Mismatches 125; Indels 131; Gaps 20;
QY 1 VSLLLAALLALA-ALAPAAAAEPLKLVILSRHGVSRPTKALMVSPDW-----PWPV 54
Db 86 IALFLPEALWMSLAGGVLSAALPLFLE-----RQGER-----RLDWLAAILPALL 131
QY 55 GLTPRGAALVILGYRYFARGLLPGCPAAGTILADVTRTGOAFA-----AGLAPGCAI 109
Db 132 GIAVALSLLLAAPLWRLVLLGPG-----LAE-----TASQAQAANLRLVLCVPL-- 178
QY 110 VHAGDDPIFHGLDTGCLDQADAILAAGEGL-----TARHGLTLAKVLPASACL- 159
Db 179 -----MLHALPSIPLQAAERFVLAGLSLLNPPVLYLALHG-QASQPEQLALACL 230
QY 160 -----ECARVVVGLGPIILASTLSE--IFILEYAOGPM 199
Db 231 GSLIMPLVLLPSLWIEGWREWHRLSGTELGEGLGGRIGPILLSNAASQGLALVERLVASL 290
QY 190 EVGWGRISAEMLLHNAQFLMNRTPYIARATPILIVTALSFA- 231
Db 291 -LGEGAVT--W-VNLARKLMN-LPLIALMSLNQVILGMMRRQGGERLALLRGLTASL 345
QY 232 VVLAGHDNTNALGGDLMLQDPDPTPPGGA-----LVFELN-----RYRVYVYQT 279
Db 346 LTLFAGVGLVAAPGLVALLPQQTAGSLPALLAWFAVPLVFGVGNALLARYAAGDT 405
QY 280 MDQLR-----NLEPILLPRILPIPG 298
Db 406 RLPLRCELLGSALNAALLAVLPLIFGLPG 434

RESULT 13
US-09-146-283-2
; Sequence 2, Application US/09146283
; Patent No. 5976546
; GENERAL INFORMATION:
; APPLICANT: Laus, Reiner
; APPLICANT: Ruegg, Curtis L.
; APPLICANT: Wu, Hongyu
; TITLE OF INVENTION: Immunostimulatory Compositions
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Denlinger & Associates
; STREET: 350 Cambridge Ave. Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/146,283

FILING DATE: 03-SEPT-1998
 CLASSIFICATION: 536
 ATTORNEY/AGENT INFORMATION:
 NAME: Judge, Linda R.
 REGISTRATION NUMBER: 42,702
 REFERENCE/DOCKET NUMBER: 7636-0010.21
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 650-324-0880
 TELEFAX: 650-324-0960
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 515 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 ORGANISM: Homo sapiens
 INDIVIDUAL ISOLATE: prostatic acid phosphatase-GM-CSF
 INDIVIDUAL ISOLATE: fusion protein; Fig. 1
 US-09-146-283-2

Query Match 6.3%; Score 103.5; DB 2; Length 515;
 Best Local Similarity 23.8%; Pred. No. 0.0096;
 Matches 92; Conservative 38; Mismatches 148; Indels 109; Gaps 20;

QY 2 SLLLAALLAALAPAAAAEPLKLVILSRHGVSPKALMVSP--DWPWPVG--L 56
 DB 13 SLSLGFLLFFWLDRLSVLAKELF--VTLVFRHGDRSPIDTPTDPIKESWPGQGL 70
 QY 57 TPRG-----AALVLLGYRYF-----ARGLLPGCPAAGTILADVDRTRTQQA 98
 DB 71 TQLGMEQHYELGEYIRKRYKFLNESYKHEQVYIRS-----TDVDRTLMSAMT 118
 QY 99 FAAGLAP--GCAI-----VHA---GDD-----PIFHGLDTGCLDQ--- 128
 DB 119 NLAAFPPEGVSINWPILLMQIPVHTVPLSEDLQLYLPFRNCPFOESETLKSEEFQ 178
 QY 129 -----ADAILAAGEGGITARHGLTLAKVLFASFACLECARVVGRLGPLLAS----- 174
 DB 179 KRLHPYKDFIATLKG--LSGLHGQDLFGIWSKVYDPLCYESVNFPLPSWATEDTWTCLR 236
 QY 175 TLSIFILLEYAGGMEVGVNGRISAEWLLHNAQFLMNTPTPIARATPILIVTALSARVVL 234
 DB 237 ELSLSLSLSLYGIHKQKSRQGGVLVNE---ILN---HMKRATQI-----PSYKKLIM 285
 QY 235 LAGHDTNLALGGLDLWQLPQDPTPPGGA-----LVFELMNRVYRVVMYQTMQDLRLNLEPL 289
 DB 286 YSAHDTTVS--GLQMALDVYNGLLPPYASCHLTLYFEKGEYFVEMY-----RNETHQ 337
 QY 290 LPRILPIPGCGSEAAACSLSDPARLVAP 316
 DB 338 EYPPLMLPGC--SPSCPLERFAELVGP 362

RESULT 15
 US-09-344-195-2
 Sequence 2, Application US/09344195
 Patent No. 6210662
 GENERAL INFORMATION:
 APPLICANT: Laus, Reiner
 APPLICANT: Wu, Hongyu
 APPLICANT: Ruegg, Curtis L.
 TITLE OF INVENTION: Immunostimulatory Compositions
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Dehlinger & Associates
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94306
 COMPUTER READABLE FORM:

FILING DATE: 03-SEPT-1998
 CLASSIFICATION: 536
 ATTORNEY/AGENT INFORMATION:
 NAME: Judge, Linda R.
 REGISTRATION NUMBER: 42,702
 REFERENCE/DOCKET NUMBER: 7636-0010.21
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 650-324-0880
 TELEFAX: 650-324-0960
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 515 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 ORGANISM: Homo sapiens
 INDIVIDUAL ISOLATE: prostatic acid phosphatase-GM-CSF
 INDIVIDUAL ISOLATE: fusion protein; Fig. 1
 US-09-146-283-2

Query Match 6.3%; Score 103.5; DB 2; Length 515;
 Best Local Similarity 23.8%; Pred. No. 0.0096;
 Matches 92; Conservative 38; Mismatches 148; Indels 109; Gaps 20;

QY 2 SLLLAALLAALAPAAAAEPLKLVILSRHGVSPKALMVSP--DWPWPVG--L 56
 DB 13 SLSLGFLLFFWLDRLSVLAKELF--VTLVFRHGDRSPIDTPTDPIKESWPGQGL 70
 QY 57 TPRG-----AALVLLGYRYF-----ARGLLPGCPAAGTILADVDRTRTQQA 98
 DB 71 TQLGMEQHYELGEYIRKRYKFLNESYKHEQVYIRS-----TDVDRTLMSAMT 118
 QY 99 FAAGLAP--GCAI-----VHA---GDD-----PIFHGLDTGCLDQ--- 128
 DB 119 NLAAFPPEGVSINWPILLMQIPVHTVPLSEDLQLYLPFRNCPFOESETLKSEEFQ 178
 QY 129 -----ADAILAAGEGGITARHGLTLAKVLFASFACLECARVVGRLGPLLAS----- 174
 DB 179 KRLHPYKDFIATLKG--LSGLHGQDLFGIWSKVYDPLCYESVNFPLPSWATEDTWTCLR 236
 QY 175 TLSIFILLEYAGGMEVGVNGRISAEWLLHNAQFLMNTPTPIARATPILIVTALSARVVL 234
 DB 237 ELSLSLSLSLYGIHKQKSRQGGVLVNE---ILN---HMKRATQI-----PSYKKLIM 285
 QY 235 LAGHDTNLALGGLDLWQLPQDPTPPGGA-----LVFELMNRVYRVVMYQTMQDLRLNLEPL 289
 DB 286 YSAHDTTVS--GLQMALDVYNGLLPPYASCHLTLYFEKGEYFVEMY-----RNETHQ 337
 QY 290 LPRILPIPGCGSEAAACSLSDPARLVAP 316
 DB 338 EYPPLMLPGC--SPSCPLERFAELVGP 362

RESULT 14
 US-08-579-823A-2
 Sequence 2, Application US/08579823A
 Patent No. 6080409
 GENERAL INFORMATION:
 APPLICANT: Laus, Reiner
 APPLICANT: Ruegg, Curtis L.
 APPLICANT: Wu, Hongyu
 APPLICANT: Dehlinger & Associates
 TITLE OF INVENTION: Immunostimulatory Composition and Method
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 ADDRESS: Dehlinger & Associates
 STREET: 350 Cambridge Ave. Suite 250
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94306
 COMPUTER READABLE FORM:

Thu May 6 08:41:38 2004

STREET: 350 Cambridge Ave. Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/344,195
FILING DATE: 24-Jun-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/145,283
FILING DATE: 03-SEPT-1998
ATTORNEY/AGENT INFORMATION:
NAME: Judge, Linda R.
REGISTRATION NUMBER: 42,702
REFERENCE/DOCKET NUMBER: 7636-0010.21
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-324-0880
TELEFAX: 650-324-0960
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 515 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
INDIVIDUAL ISOLATE: Prostatic acid phosphatase-GM-CSF
fusion protein; Fig. 1
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-344-195-2

Query Match 6.3%; Score 103.5; DB 3; Length 515;
Best Local Similarity 23.8%; Pred. No. 0.0096;
Matches 92; Conservative 38; Mismatches 148; Indels 109; Gaps 20;
QY 2 SLIAALLAALAPAAAAEPLEKWLGRHVSPTKALMVSP--DWFPVVG--L 56
Db 13 SLISGLFLFLFFLDRLSLAKELKF--VTLVFEHGRSPIDTFPTPIKESWPGQGL 70
QY 57 TPRG-----AALVLGYRYP-----ARGLLPGCPAAGTILADVDRTRTGOA 98
Db 71 TQLGMEQHYELGEYIRKRYKFLNESYKHEQVYIRS-----TDVDTILMSAMT 118
QY 99 FAAGLAP--GCAJ-----VHA--GDD-----PIFHGLDTGCLDQ---- 128
Db 119 NLAAFPPEGVSTWNPILLWQIPVHTVPLSEDLILYLPFRNCFRQEESETLKSEETQ 178
QY 129 -----ADAILAGEGLTARHGLTAKVLFASFACLECARVVGRLGPLLAS----- 174
Db 179 KRLHPYKDFATLTK--LSGLHQDLFGIMSKVYDPLYCESVHNFTLPSWATEDTMTKL 236
QY 175 TLSEIFLEVAQGMWGVGRISAEMLHNAQFLMNRTPVIAEATPILIVTALS PARVVL 234
Db 237 ELSELSLSLYGTHKQEKSRLOGGVLIWE---ILN---HMKRATQI-----PSYKKLIM 285
QY 235 LAGHDTNLALGGDLMLQPDQDETPPGA-----LVFELMNRVYRVYQTMQLRNLLEPL 289
Db 286 YSAHDTTVS--GLQMALDVYNGLLPPYASCHLTLYFEKGEYFVEMY-----RNETHQ 337
QY 290 LPRILPIPGGSEACSLSDPARIVAP 316
Db 338 EPYFLMLPGC--SPSCPLERFAELVGP 362

Search completed: April 30, 2004, 12:43:19
Job time : 14.7422 secs

GenCore version 5.1.1.6
 Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 30, 2004, 12:41:31 ; Search time 31.0244 Seconds
 (without alignments)
 2841.193 Million cell updates/sec

Title: US-10-021-723B-14
 Perfect score: 1647
 Sequence: 1 VSLIAALLAALAPAAAA.....CGSEAAACSLSPARLVAPAC 318

Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 1138120 seqs, 277189581 residues

Total number of hits satisfying chosen parameters: 1138120

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database :				Published Applications AA:*			
1:	/cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep:*						
2:	/cgn2_6/prodata/1/pubpaa/CTC_NEW_PUB.pep:*						
3:	/cgn2_6/prodata/1/pubpaa/US03_NEW_PUB.pep:*						
4:	/cgn2_6/prodata/1/pubpaa/US06_PUBCOMB.pep:*						
5:	/cgn2_6/prodata/1/pubpaa/US07_NEW_PUB.pep:*						
6:	/cgn2_6/prodata/1/pubpaa/CTC_PUBCOMB.pep:*						
7:	/cgn2_6/prodata/1/pubpaa/US08_NEW_PUB.pep:*						
8:	/cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pep:*						
9:	/cgn2_6/prodata/1/pubpaa/US09C_PUBCOMB.pep:*						
10:	/cgn2_6/prodata/1/pubpaa/US09C_PUBCOMB.pep:*						
11:	/cgn2_6/prodata/1/pubpaa/US09C_PUBCOMB.pep:*						
12:	/cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pep:*						
13:	/cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pep:*						
14:	/cgn2_6/prodata/1/pubpaa/US10C_PUBCOMB.pep:*						
15:	/cgn2_6/prodata/1/pubpaa/US10C_PUBCOMB.pep:*						
16:	/cgn2_6/prodata/1/pubpaa/US10C_PUBCOMB.pep:*						
17:	/cgn2_6/prodata/1/pubpaa/US60_NEW_PUB.pep:*						
18:	/cgn2_6/prodata/1/pubpaa/US60_PUBCOMB.pep:*						

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1647	100.0	318	14	US-10-021-723A-14
2	663.5	40.3	409	14	US-10-021-723A-8
3	650	39.5	421	14	US-10-021-723A-6
4	632.5	38.4	430	9	US-09-866-379-10
5	627.5	38.1	436	14	US-10-156-660-2
6	616.5	37.4	412	14	US-10-334-672-1
7	616.5	37.4	412	14	US-10-334-671-1
8	613.5	37.2	410	14	US-10-021-723A-13
9	613.5	37.2	410	14	US-10-021-723A-15
10	613.5	37.2	432	9	US-09-866-379-8
11	613.5	37.2	432	12	US-10-282-122A-43351
12	613.5	37.2	432	14	US-10-156-660-4
13	611.5	37.1	440	9	US-09-777-566A-2
14	611.5	37.1	440	9	US-09-866-379-2
15	611.5	37.1	440	14	US-10-034-985-2

16	611.5	37.1	440	15	US-10-430-356-2	Sequence 2, Appli
17	611	37.1	476	14	US-10-021-723A-10	Sequence 10, Appli
18	608.5	36.9	432	15	US-10-284-962-3	Sequence 3, Appli
19	608.5	36.9	432	15	US-10-284-962-14	Sequence 14, Appli
20	608.5	36.9	433	14	US-10-266-041-1	Sequence 1, Appli
21	602.5	36.6	432	15	US-10-284-962-5	Sequence 5, Appli
22	596	36.2	441	14	US-10-021-723A-12	Sequence 12, Appli
23	590	35.8	441	12	US-10-282-122A-77792	Sequence 77792, A
24	589	35.8	441	14	US-10-021-723A-2	Sequence 2, Appli
25	567	34.4	420	14	US-10-021-723A-4	Sequence 4, Appli
26	554.5	33.7	432	14	US-10-021-723A-16	Sequence 16, Appli
27	138	8.4	426	12	US-10-257-174-44	Sequence 44, Appli
28	138	8.4	426	12	US-10-343-357-6	Sequence 6, Appli
29	117	7.1	502	12	US-10-389-647-684	Sequence 684, App
30	109.5	6.6	3668	15	US-10-402-842-4	Sequence 4, Appli
31	108	6.6	401	15	US-10-369-493-20570	Sequence 20570, A
32	107.5	6.5	801	9	US-09-822-827-982	Sequence 982, App
33	107.5	6.5	801	9	US-09-855-793-982	Sequence 982, App
34	104	6.3	474	16	US-10-354-437-32	Sequence 32, Appli
35	103.5	6.3	386	9	US-09-822-827-945	Sequence 945, App
36	103.5	6.3	386	9	US-09-895-793-945	Sequence 945, App
37	103.5	6.3	386	10	US-09-974-546-48	Sequence 48, Appli
38	103.5	6.3	386	15	US-10-341-434-214	Sequence 214, App
39	102.5	6.2	4551	10	US-09-793-708-1	Sequence 1, Appli
40	102.5	6.2	4551	14	US-10-201-365-2	Sequence 2, Appli
41	102.5	6.2	4551	14	US-10-160-539-1	Sequence 1, Appli
42	102.5	6.2	4613	9	US-09-861-289-31	Sequence 31, Appli
43	102.5	6.2	4613	9	US-09-860-846-31	Sequence 31, Appli
44	102.5	6.2	4613	10	US-09-988-384B-31	Sequence 31, Appli
45	102.5	6.2	4613	10	US-09-836-821-31	Sequence 31, Appli

ALIGNMENTS

RESULT 1
 US-10-021-723A-14
 ; Sequence 14, Application US/10021723A
 ; Publication No. US20030101476A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Short, Jay
 ; APPLICANT: Mathur, Eric
 ; APPLICANT: Richardson, Toby
 ; APPLICANT: Robertson, Dan
 ; APPLICANT: Barton, Nelson
 ; TITLE OF INVENTION: Recombinant Phytases and Uses Thereof
 ; FILE REFERENCE: 112766.140 (DIV-016C1P)
 ; CURRENT APPLICATION NUMBER: US/10/021,723A
 ; CURRENT FILING DATE: 2002-10-23
 ; PRIOR APPLICATION NUMBER: US 60/255,090
 ; PRIOR FILING DATE: 2000-12-12
 ; NUMBER OF SEQ ID NOS: 74
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 14
 ; LENGTH: 318
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: consensus sequence
 US-10-021-723A-14

Query Match 100.0%; Score 1647; DB 14; Length 318;
 Best Local Similarity 100.0%; Pred. No. 66-149;
 Matches 318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSLIAALLAALAPAAAAEPLKLEKVVILSRHGVSRPTKALMVSPDPWFWVGLTPRG 60

DB 1 VSLIAALLAALAPAAAAEPLKLEKVVILSRHGVSRPTKALMVSPDPWFWVGLTPRG 60

QY 61 AALVLLGYRFARGLLPGCPAAGTILADVDERTRTGTQAFAGLAPGCAIVHAGDDPIFHG 120

DB 61 AALVLLGYRFARGLLPGCPAAGTILADVDERTRTGTQAFAGLAPGCAIVHAGDDPIFHG 120


```
QY 121 LDTGCLDQADAILAAGEGLTARHGLTLAKVNFASACLECARVVGRLGPILASTLSEIF 180
Db 121 LDTGCLDQADAILAAGEGLTARHGLTLAKVNFASACLECARVVGRLGPILASTLSEIF 180
QY 181 LLEYAQGMVEGWGRISAEWLLHNAQFLMNRTPYIARATPILIVTALSAPARVVLGAGDT 240
Db 181 LLEYAQGMVEGWGRISAEWLLHNAQFLMNRTPYIARATPILIVTALSAPARVVLGAGDT 240
QY 241 NLALGGLDLWOLPQDETPPGGALVFELWNRVVRVMYQTMQDLNLELLPRILPIGCG 300
Db 241 NLALGGLDLWOLPQDETPPGGALVFELWNRVVRVMYQTMQDLNLELLPRILPIGCG 300
QY 301 SEACSLSDPARLVAPAC 318
Db 301 SEACSLSDPARLVAPAC 318

RESULT 2
US-10-021-723A-8
; Sequence 8, Application US/10021723A
; Publication No. US20030101476A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Mathur, Eric
; APPLICANT: Richardson, Toby
; APPLICANT: Robertson, Dan
; APPLICANT: Barton, Nelson
; TITLE OF INVENTION: Recombinant Phytases and Uses Thereof
; FILE REFERENCE: 112766.140 (DIV-016CIP)
; CURRENT APPLICATION NUMBER: US/10/021,723A
; CURRENT FILING DATE: 2002-10-23
; PRIOR APPLICATION NUMBER: US 60/255,090
; PRIOR FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Rhizobium
US-10-021-723A-8

Query Match 40.3%; Score 663.5; DB 14; Length 409;
Best Local Similarity 54.0%; Pred. No. 8.8e-55;
Matches 218; Conservative 22; Mismatches 75; Indels 89; Gaps 39;

QY 2 SILL-AALLA-LAALAPAA-----AAAEPLKLEKVVILSRHGVSPPTKALMVSP-----DW 50
Db 4 SILLPALLAGCAASAPACASPAAPGSLKLEKVVMLMRHGVPRPTKAAVVPVGSATW 63
QY 51 P-WPVG---LTPRGAALV-LIGYR---YF-ARGLLP-GCPAAGTIL--ADVDERT-RTGQ 97
Db 64 PDWPVDFGLLTPHGAAGVKLLGESDRLYFGGRGLFPFGCPAAGTIVLKASYKERTIATAQ 123
QY 98 AFAAGLAPGCA--IVH-AG--DDPIFHGLDTGCLD-----QADAILA-AGEGGLTA---- 142
Db 124 SWAGFMFGCTADVAHPAGPDDDAIFHGLDGGPASFDGKRAFDAAQAPEGGLTASTAR 183
QY 143 -RHGLT-LAKVNFSA-SAC---LECARVVGSR-----LGPL-LASTLSIFLLEYAQG- 187
Db 184 HRGELTLAKVNLNCPALPACPLIAPESRLVAQPHDRPEIEGLDVGSTASQTLVLEYEGK 243
QY 188 PM-EVGMGRIS---AEWLL--HNAQF-LMNRTPYI--RATPIL--IVTAL---SPARVV 233
Db 244 PMAEVGMGRVSRABIEQLLRHPLKFRYSNRPGYIAATAAAPIVREIVTALGDRSPARLT 303
QY 234 LLAGHDTNLA-LGG-LDL-WOLPQ--PDETPPGGALVFEL-----WNRVVRVMY--OTMD 281
Db 304 LLAGHDTNVDLGGFFDLHWQVPSYPADEVPPGSGALGFELVSNKAGDRYVRAFIRGQTM 363
QY 282 QLRNLEP-----LLPRILPIPGCGSE---AACSLSDPARLVAP 316
Db 364 QLRNLEPRLSDTLFRYRLPIPGCGNSVEATACAWSDFARLAAP 407
```

RESULT 3

```
US-10-021-723A-6
; Sequence 6, Application US/10021723A
; Publication No. US20030101476A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Mathur, Eric
; APPLICANT: Richardson, Toby
; APPLICANT: Robertson, Dan
; APPLICANT: Barton, Nelson
; TITLE OF INVENTION: Recombinant Phytases and Uses Thereof
; FILE REFERENCE: 112766.140 (DIV-016CIP)
; CURRENT APPLICATION NUMBER: US/10/021,723A
; CURRENT FILING DATE: 2002-10-23
; PRIOR APPLICATION NUMBER: US 60/255,090
; PRIOR FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: 953-6 phytase sequence
US-10-021-723A-6
```

```
Query Match 39.5%; Score 650; DB 14; Length 421;
Best Local Similarity 53.0%; Pred. No. 1.8e-53;
Matches 213; Conservative 20; Mismatches 81; Indels 88; Gaps 37;

QY 3 LLLAALLA-LAALAPAA-----AAAEPLKLEKVVILSRHGVSPPTKALMVSP-----DWP- 51
Db 18 LLSAALLAGCAASAPACASPAAPGSLKLEKVVMLMRHGVPRPTKAAVVPVGSATWPD 77
QY 52 WPVG---LTPRGAALV-LIG---YRFARGLLP-GCPAAGTIL--ADVDERT-RTGQAF 99
Db 78 WPVDFGLLTPHGAAGVKLLGESDRLTFGGRGLFPDGCFAAGTIVLKASYKERTIATAQNW 137
QY 100 AAGLAPGCA--IVH-AG--DDPIFHGLDTGCLD-----QADAILA-AGEGGLTA----R 143
Db 138 AAGFMFGCTADVAHPAGPDDDAIFHGLDGGPASFDGKRAFDAAQAPEGGLTASTARHR 197
QY 144 HGLT-LAKVNFSA-SAC---LECARVVGSR-----LGPL-LASTLSIFLLEYAQG-PM 189
Db 198 GELTLAKVNLNCPALPACPLIAPESRLVAQPHDRPEIEGLDVGSTASQTLVLEYEGKPM 257
QY 190 -EVGMGRIS---AEWLL--HNAQF-LMNRTPYI--ARATPIL--IVTAL---SPARVVLL 235
Db 258 AEVGMGRVSRABIEQLLRHPLKFRYSNRPGYIAATAAAPIVREIVTALGDRSPARLTIL 317
QY 236 AGHDTNLA-LGG-LDL-WOLPQ--PDETPPGGALVFEL-----WNRVVRVMY--OTMDOL 283
Db 318 AGHDTNVDLGGFFDLHWQVPSYPADEVPPGSGALGFELVSNKAGDRYVRAFIRGQTM 377
QY 284 RNLEP-----LLPRILPIPGCG---SEACSLSDPARLVAP 316
Db 378 RNLEPLGSDALYRRLPIPGCGHSVEATACAWSDFARLAAP 419
```

RESULT 4

```
US-09-866-379-10
; Sequence 10, Application US/09866379
; Patent No. US20020136754A1
; GENERAL INFORMATION:
; APPLICANT: DIVERSA CORPORATION
; APPLICANT: Short, Jay
; APPLICANT: KRETZ, Keith
; APPLICANT: GRAY, Kevin
; APPLICANT: BARTON, Nelson
; APPLICANT: GARRETT, James
; APPLICANT: O'DONOGHUE, Eileen
; TITLE OF INVENTION: RECOMBINANT BACTERIAL PHYTASES AND USES THEREOF
```

```
; FILE REFERENCE: DIVER1370-7
; CURRENT APPLICATION NUMBER: US/09/866,379
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 09/580,515
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 09/318,528
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: US 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: US 08/910,798
; PRIOR FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Modified phytase
US-09-866-379-10

Query Match      38.4%; Score 632.5; DB 9; Length 430;
Best Local Similarity 45.2%; Pred. No. 8.6e-52;
Matches 194; Conservative 44; Mismatches 76; Indels 115; Gaps 35;

QY  2 SLLLAALLAALAP--AAAAAP-LKLEKVVILSRHGVSRPTKA--LM--VSPD-WP-W 52
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  3 AILIPFLSLILPTPQSAFAQSEPELKLESVIVSRHGVRAFTKATQLMQDVTPDWPTW 62

QY  53 PV---GLTPRGAALV-LLGY-----RYFARGLLP--GCPAAG--TILADVDERTR-TGOAP 99
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  63 PVKLGELTPRGELIAYLGHYWRQLVADGLLPKCGCPQSGQVAILADVDERTRKTGEAF 122

QY  100 AAGLAPCAI-VH-----AGDDPIFGHLDTG-C-LDOA---DAILAAGEGLTARHG--- 145
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  123 AAGLAPCAITVHTQADTSSPDPLFNPPLKTVGCOLDNANVTDAILEAG-GSIADFTGHYQ 181

QY  146 ---LTLAKVNF--ASACL-----ECARVVGRLGPELLASTLS 177
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  182 TAFRELERVLPQSNLCKREKQDECSLTQALPSELKVSADCVSLTGAVS--LASMLT 239

QY  178 EIFLELYAOGPMVEGWGRI--SAEW---LLHNAQFLMNETPYIA--RATPI--LIVTA 227
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  240 EIFLQQAQGMPEPGWGRITDSHQWNTLLSLHNAQFOLLQRTPEVARSRATPLDLIKTA 299

QY  228 S-----PARVVLLAGHDTNLA-LGG-LDL-WOLP--QPDETTPGGALVFELWN 270
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  300 TEPHPQQAQGVTLPTSFLFIAGHDTNLANLGGALELNLWTLPGQPDNTPGGELVFERWR 359

QY  271 R-----YVRVMYQTMQDLRLNLEPLPRILP-----IPGC---GSEACSLSDFARL 313
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  360 RLSDNSQWIQVSLVFQTLQQRDKTPLUSLNTPPGEVKLTLAGCEERNAQMGCSLAGFTQI 419

QY  314 V----APAC 318
   : : : : :
Db  420 VNEARIEAC 428

RESULT 5
US-10-156-660-2
; Sequence 2, Application US/10156660
; Publication No. US20030103958A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay M.
; APPLICANT: Kretz, Keith
; APPLICANT: Gray, Kevin A.
; APPLICANT: Barton, Nelson R.
; APPLICANT: Garrett, James B.
; APPLICANT: O'Donoghue, Eileen
; APPLICANT: Mathur, Eric J.
; TITLE OF INVENTION: PHYTASES, NUCLEIC ACIDS ENCODING THEM
```

```
; TITLE OF INVENTION: AND METHODS FOR MAKING AND USING THEM
; FILE REFERENCE: 09010-029007
; CURRENT APPLICATION NUMBER: US/10/156,660
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: US 09/866,379
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 09/580,515
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 09/318,528
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: US 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: US 09/259,214
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: US 08/910,798
; PRIOR FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 436
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: modified phytase enzyme
US-10-156-660-2

Query Match      38.1%; Score 627.5; DB 14; Length 436;
Best Local Similarity 44.9%; Pred. No. 2.6e-51;
Matches 193; Conservative 44; Mismatches 78; Indels 115; Gaps 35;

QY  2 SLLLAALLAALAP--AAAAAP-LKLEKVVILSRHGVSRPTKA--LM--VSPD-WP-W 52
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  3 AILIPFLSLILPTPQSAFAQSEPELKLESVIVSRHGVRAFTKATQLMQDVTPDWPTW 62

QY  53 PV---GLTPRGAALV-LLGY-----RYFARGLLP--GCPAAG--TILADVDERTR-TGOAP 99
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  63 PVKLGELTPRGELIAYLGHYWRQLVADGLLPKCGCPQSGQVAILADVDERTRKTGEAF 122

QY  100 AAGLAPCAI-VH-----AGDDPIFGHLDTG-C-LDOA---DAILAAGEGLTARHG--- 145
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  123 AAGLAPCAITVHTQADTSSPDPLFNPPLKTVGCOLDNANVTDAILEAGGSIADFTGHYQ 182

QY  146 ---LTLAKVNF--ASACL-----ECARVVGRLGPELLASTLS 177
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  183 TAFRELERVLPQSNLCKREKQDECSLTQALPSELKVSADCVSLTGAVS--LASMLT 240

QY  178 EIFLELYAOGPMVEGWGRI--SAEW---LLHNAQFLMNETPYIA--RATPI--LIVTA 226
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  241 EIFLQQAQGMPEPGWGRITDSHQWNTLLSLHNAQFOLLQRTPEVARSRATPLDLIKTA 300

QY  227 LS-----PARVVLLAGHDTNLA-LGG-LDL-WOLP--QPDETTPGGALVFELW 269
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  301 LTPHPQQAQGVTLPTSFLFIAGHDTNLANLGGALELNLWTLPGQPDNTPGGELVFERW 360

QY  270 NR-----YVRVMYQTMQDLRLNLEPLPRILP-----IPGC---GSEACSLSDFAR 312
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  361 RLSDNSQWIQVSLVFQTLQQRDKTPLUSLNTPPGEVKLTLAGCEERNAQMGCSLAGFTQ 420

QY  313 LV----APAC 318
   : : : : :
Db  421 IVNEARIEAC 430

RESULT 6
US-10-334-672-1
; Sequence 1, Application US/10334672
; Publication No. US20030157646A1
; GENERAL INFORMATION:
; APPLICANT: Lanahan, Mike
; APPLICANT: Koepf, Edward
; APPLICANT: Kretz, Keith
; TITLE OF INVENTION: Microbially-Expressed Thermotolerant Phytase For Animal Feed
; FILE REFERENCE: SYNG-P01-001
```

```
/ CURRENT APPLICATION NUMBER: US/10/334,672
/ CURRENT FILING DATE: 2002-12-30
/ PRIOR APPLICATION NUMBER: 60/344,523
/ PRIOR FILING DATE: 2001-12-28
/ NUMBER OF SEQ ID NOS: 4
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 1
/ LENGTH: 412
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: No. US20030157646A19X Phytase
US-10-334-672-1

Query Match      37.4%; Score 616.5; DB 14; Length 412;
Best Local Similarity 45.5%; Pred. No. 2.8e-50;
Matches 187; Conservative 41; Mismatches 70; Indels 113; Gaps 34;

QY 19 AAAP-LKLEKVVILSRHGVRSPTKA--LM--VSPD-WP-WPV---GLTPRGAALV-LIG 67
Db 2 AQSEPELKLESVIVSRHGVRAPTKATQMQDVTDPDAMPTVPVKLGELTPRGGLIAYLG 61

QY 68 Y----RYFARGLLP--GCPAAG--TILADVDRTR--TQAFAGAGLAPGCAI--VH-----A 112
Db 62 HYWRQLVADGLLPKCGCPQSGQVAIIADVDRTRTKTGEAFAGLAPDCAITVHTQADTS 121

QY 113 GDDPIFHGLDTG-C-LDQA---DAILAGEGGLTARHG-----LTLAKVLNF--ASACL 159
Db 122 SPDPLFNPLKTVGCQLDNNVTDAILERAGGSIADFTGHYQTAFARELERVLPFQSNLCL 181

QY 160 -----ECARVVGRIGLPILLASTLSLSEIFLLEYAAGPMEVWGRI 196
Db 182 KREKQDESCSLTQALPSELKVSADCVSLTGAVS--LASMLTEIFLLQQAQGMPEFGWRI 239

QY 197 --SAEW---LHNAQF-LMNRTPYIA--RATPI--LIVTALS-----PARV 232
Db 240 TDSHOWNTLLSHNAQFDLLQRTPEVARSRATPLDLIKTALTTPHPQQAQGVLTPTSV 299

QY 233 VLLAGHDTNLA-LGG-LDL-WOLP-QPDETPPGGALVPFELMNR-----YVRVMYQTM 280
Db 300 LFIAGHDTNLNLGGALELNWTLPGQDNTTPGGELVFERWRLSDNSQWIVQSLVFQTL 359

QY 281 DQIRNLEPLLPRILP-----IPGC---GSEAAACSLSDPARLV-----APAC 318
Db 360 QQMRDKTPLSLNTPPGEVKLTLAGCERNQAQMCSLAGFTQIVNEARIIPAC 410

RESULT 7
US-10-334-671-1
/ Sequence 1, Application US/10334671
/ Publication No. US20030170293A1
/ GENERAL INFORMATION:
/ APPLICANT: Lanahan, Mike
/ APPLICANT: Koepf, Edward
/ APPLICANT: Kretz, Keith
/ TITLE OF INVENTION: Microbially-Expressed Thermotolerant Phytase For Animal Feed
/ FILE REFERENCE: 70098
/ CURRENT APPLICATION NUMBER: US/10/334,671
/ CURRENT FILING DATE: 2002-12-30
/ PRIOR APPLICATION NUMBER: 60/344,523
/ PRIOR FILING DATE: 2001-12-28
/ NUMBER OF SEQ ID NOS: 4
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 1
/ LENGTH: 412
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: No. US20030170293A19X Phytase
US-10-334-671-1

Query Match      37.4%; Score 616.5; DB 14; Length 412;
Best Local Similarity 45.5%; Pred. No. 2.8e-50;
```

```
Matches 187; Conservative 41; Mismatches 70; Indels 113; Gaps 34;

QY 19 AAAP-LKLEKVVILSRHGVRSPTKA--LM--VSPD-WP-WPV---GLTPRGAALV-LIG 67
Db 2 AQSEPELKLESVIVSRHGVRAPTKATQMQDVTDPDAMPTVPVKLGELTPRGGLIAYLG 61

QY 68 Y----RYFARGLLP--GCPAAG--TILADVDRTR--TQAFAGAGLAPGCAI--VH-----A 112
Db 62 HYWRQLVADGLLPKCGCPQSGQVAIIADVDRTRTKTGEAFAGLAPDCAITVHTQADTS 121

QY 113 GDDPIFHGLDTG-C-LDQA---DAILAGEGGLTARHG-----LTLAKVLNF--ASACL 159
Db 122 SPDPLFNPLKTVGCQLDNNVTDAILERAGGSIADFTGHYQTAFARELERVLPFQSNLCL 181

QY 160 -----ECARVVGRIGLPILLASTLSLSEIFLLEYAAGPMEVWGRI 196
Db 182 KREKQDESCSLTQALPSELKVSADCVSLTGAVS--LASMLTEIFLLQQAQGMPEFGWRI 239

QY 197 --SAEW---LHNAQF-LMNRTPYIA--RATPI--LIVTALS-----PARV 232
Db 240 TDSHOWNTLLSHNAQFDLLQRTPEVARSRATPLDLIKTALTTPHPQQAQGVLTPTSV 299

QY 233 VLLAGHDTNLA-LGG-LDL-WOLP-QPDETPPGGALVPFELMNR-----YVRVMYQTM 280
Db 300 LFIAGHDTNLNLGGALELNWTLPGQDNTTPGGELVFERWRLSDNSQWIVQSLVFQTL 359

QY 281 DQIRNLEPLLPRILP-----IPGC---GSEAAACSLSDPARLV-----APAC 318
Db 360 QQMRDKTPLSLNTPPGEVKLTLAGCERNQAQMCSLAGFTQIVNEARIIPAC 410

RESULT 8
US-10-021-723A-13
/ Sequence 13, Application US/10021723A
/ Publication No. US20030101476A1
/ GENERAL INFORMATION:
/ APPLICANT: Short, Jay
/ APPLICANT: Richardson, Toby
/ APPLICANT: Robertson, Dan
/ APPLICANT: Barton, Nelson
/ TITLE OF INVENTION: Recombinant Phytases and Uses Thereof
/ FILE REFERENCE: 112766.140 (DIV-016CIP)
/ CURRENT APPLICATION NUMBER: US/10/021,723A
/ CURRENT FILING DATE: 2002-10-23
/ PRIOR APPLICATION NUMBER: US 60/255,090
/ PRIOR FILING DATE: 2000-12-12
/ NUMBER OF SEQ ID NOS: 74
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 13
/ LENGTH: 410
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: amino acid sequence of a phytase enzyme with
/ OTHER INFORMATION: predicted glycosylation sites
US-10-021-723A-13

Query Match      37.2%; Score 613.5; DB 14; Length 410;
Best Local Similarity 45.5%; Pred. No. 5.3e-50;
Matches 186; Conservative 41; Mismatches 69; Indels 113; Gaps 34;

QY 21 AEP-LKLEKVVILSRHGVRSPTKA--LM--VSPD-WP-WPV---GLTPRGAALV-LIGY- 68
Db 2 SEPELKLESVIVSRHGVRAPTKATQMQDVTDPDAMPTVPVKLGELTPRGGLIAYLGHY 61

QY 69 ---RYFARGLLP--GCPAAG--TILADVDRTR--TQAFAGAGLAPGCAI--VH-----AGD 114
Db 62 WRQRLVADGLLPKCGCPQSGQVAIIADVDRTRTKTGEAFAGLAPDCAITVHTQADTSSP 121

QY 115 DPFPHGLDTG-C-LDQA---DAILAGEGGLTARHG-----LTLAKVLNF--ASACL-- 159
Db 122 DPLFNPLKTVGCQLDNNVTDAILERAGGSIADFTGHYQTAFARELERVLPFQSNLCLR 181
```

QY 160 -----ECARVVGRGLPGLASTLGEIFLLEVAQGMVGVGRI-- 196
Db 182 EKODESCLTQALPSELKVSADCVSLTGAVS--LASMTEIFLLQQAQGPFGWGRIND 239
QY 197 SAEW---LLHNAQF-LMNRTPYIA--RATPI--LIVTALS-----PARVVL 234
Db 240 SHQWNTLLSLHNAQFDLLQRTPEVARSRAATPLDLIKTALTTPHPKQAYGVTLPTSVLF 299
QY 235 LAGHDTNLA-LGG-LDL-WQLP--QPDETPGGALVFELWNR-----YVRVMYQTMDO 282
Db 300 IAGHDTNLANLGGALELWTLPGQPDNTPPGGELVFERWRLSDNSQWISQVSLVFTLQ 359
QY 283 LRNLEPLPRILP-----IPGC---GSEAACSLSDFARLV-----APAC 318
Db 360 MRDKTFLSLNTPPGVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPAC 408
RESULT 9
US-10-021-723A-15
; Sequence 15, Application US/10021723A
; Publication No. US20030101476A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Mathur, Eric
; APPLICANT: Richardson, Toby
; APPLICANT: Robertson, Dan
; APPLICANT: Barton, Nelson
; TITLE OF INVENTION: Recombinant Phytases and Uses Thereof
; FILE REFERENCE: 112766.140 (DIV-016CIP)
; CURRENT APPLICATION NUMBER: US/10/021,723A
; CURRENT FILING DATE: 2002-10-23
; PRIOR APPLICATION NUMBER: US 60/255,090
; PRIOR FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 410
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: amino acid sequence of a phytase enzyme with
; OTHER INFORMATION: predicted glycosylation sites
US-10-021-723A-15
Query Match 37.2%; Score 613.5; DB 14; Length 410;
Best Local Similarity 45.5%; Pred. No. 5.3e-50;
Matches 186; Conservative 41; Mismatches 69; Indels 113; Gaps 34;
QY 21 AEP-LKLEKVVILSRHGVRSPTKA--LM--VSPD-WP-WPV---GLTPRGAALV-LLGY- 68
Db 2 SEPELKLESVVIVSRHGVRAPTKATQLMQDVTDPAMPTWPKLGLTTPRGGLIAYLGHV 61
QY 69 ---RYFARGILP--GCPAAG--TILADVDERTR--TGAFAAGLAPGCAI-VH-----AGD 114
Db 62 WRORLVADGLLPKGCPCQSQGVAILADVDERTKATGEAFAAGLAPDCAITVHTQADTSSP 121
QY 115 DPFPHGLDTG-C-LDQA---DAILAAGEGGLTARHG-----LTLAKVLNF--ASACL-- 159
Db 122 DPLFNPPLKTVGCQDLDNANVTDAILERAGGSIADFTGHYQTAFRELEVLNFPQSNLCLXR 181
QY 160 -----ECARVVGRGLPGLASTLGEIFLLEVAQGMVGVGRI-- 196
Db 182 EKODESCLTQALPSELKVSADCVSLTGAVS--LASMTEIFLLQQAQGPFGWGRIND 239
QY 197 SAEW---LLHNAQF-LMNRTPYIA--RATPI--LIVTALS-----PARVVL 234
Db 240 SHQWNTLLSLHNAQFDLLQRTPEVARSRAATPLDLIKTALTTPHPKQAYGVTLPTSVLF 299
QY 235 LAGHDTNLA-LGG-LDL-WQLP--QPDETPGGALVFELWNR-----YVRVMYQTMDO 282
Db 300 IAGHDTNLANLGGALELWTLPGQPDNTPPGGELVFERWRLSDNSQWISQVSLVFTLQ 359

QY 283 LRNLEPLPRILP-----IPGC---GSEAACSLSDFARLV-----APAC 318
Db 360 MRDKTFLSLNTPPGVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPAC 408
RESULT 10
US-09-866-379-8
; Sequence 8, Application US/09866379
; Patent No. US20020136754A1
; GENERAL INFORMATION:
; APPLICANT: DIVERSA CORPORATION
; APPLICANT: Short, Jay
; APPLICANT: Kretz, Keith
; APPLICANT: Gray, Kevin
; APPLICANT: Barton, Nelson
; APPLICANT: Garrett, James
; APPLICANT: O'Donoghue, Eileen
; TITLE OF INVENTION: RECOMBINANT BACTERIAL PHYTASES AND USES THEREOF
; FILE REFERENCE: DIVER1370-7
; CURRENT APPLICATION NUMBER: US/09/866,379
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 09/580,515
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 09/318,528
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: US 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: US 09/259,214
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: US 08/910,798
; PRIOR FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 8
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-866-379-8
Query Match 37.2%; Score 613.5; DB 9; Length 432;
Best Local Similarity 45.3%; Pred. No. 5.7e-50;
Matches 194; Conservative 44; Mismatches 79; Indels 111; Gaps 36;
QY 2 SLLLAALLAALAP--AAAAAEP-LKLEKVVILSRHGVRSPTKA--LM--VSPD-WP-W 52
Db 3 ALLPPLSLIPLTPQSAFAQSEPELKLESVVIVSRHGVRAPTKATQLMQDVTDPAMPTW 62
QY 53 PV---GLTPRGAALV-LLGY---RYFARGILP--GCPAAG--TILADVDERTR--TGA 99
Db 63 PVKGLWLPTRGGELIAYLGHYQRTORLVADGLLAKKGCPCQSQGVAILADVDERTKIG 122
QY 100 AAGLAPGCAI-VH-----AGDDPIFPHGLDTG-C-LDQA---DAILAAGEGGLT--ARH 146
Db 123 AAGLAPDCAITVHTQADTSSPDPLFNPPLKTVGCQDLDNANVTDAILSRAGGSIADFT 182
QY 147 T---LAKVLNF--ASACLE-----CARVVGRGLP-LASTLSEI 179
Db 183 TAFRELEVLNFPQSNLCLREKODESCLTQALPSELKVSADNVSUVTGAVSLASMLTEI 242
QY 180 FLLEVAQGMVGVGRI--SAEW---LLHNAQF-LMNRTPYIA--RATPI--LIVTALS 228
Db 243 FLLOQAQGMPEFGWGRITDQWNTLLSLHNAQFVLLQRTPEVARSRAATPLDLIKTALT 302
QY 229 -----PARVVLLAGHDTNLA-LGG-LDL-WQLP--QPDETPGGALVFELWNR 271
Db 303 PHPPKQAYGVTLPTSVLFIAGHDTNLANLGGALELWTLPGQPDNTPPGGELVFERWR 362
QY 272 -----YVRVMYQTMDOLENLEPLPRILP-----IPGC---GSEAACSLSDFAR 314
Db 363 LSDNSQWISQVSLVFTLQQAQGMVGVGRI--LAKVLNF--ASACLE-----CARVVGRGLP-LASTLSEI 422
QY 315 -----APAC 318
|||

```
Db 423 NEARIPAC 430

RESULT 11
US-10-282-122A-43351
; Sequence 43351, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELIURA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 43351
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-282-122A-43351

Query Match 37.2%; Score 613.5; DB 12; Length 432;
Best Local Similarity 45.3%; Pred. No. 5.7e-50;
Matches 194; Conservative 44; Mismatches 79; Indels 111; Gaps 36;

QY 2 SLILAAALAAALAP--AAAAAP-LKLEKVVILSRHGVSPTKA--LM--VSPD-WP-W 52
Db 3 AILIPFSLILPTPOSAPFAQSEPELKESVIVSRHGVRAPTKATQLMQDVTPDAPTW 62
QY 53 PV---GLTPRGAALV-LLGY-----RYFARGLL--PGCPAAG--TILADVDETR-TGOAF 99
Db 63 PVKLGWLTTPRGGLIAYLGHYQRLVADGLLAKKGCPSQGVAVIADVDERTKRTGEAF 122
QY 100 AAGLAPGCAT-VH-----AGDDPIFHGLDTG-C-LDOA---DAILAAGEGLT--ARHGL 146
Db 123 AAGLAPDCAITVHTQADTSSPDLFNLKTVGVCLDNANVTDAILSRAGGSADFTGHRQ 182
QY 147 T-----LAKVLNF--ASACLE-----CARVVGRLGFL--LASTLSEI 179
Db 183 TAFRELERVLNFPQSNLCKREKQDESCSLTQALPSKLVSDNVSLTGAVSLASMLTEI 242

RESULT 12
US-10-156-660-4
; Sequence 4, Application US/10156660
; Publication No. US20030103958A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay M.
; APPLICANT: Kretz, Keith
; APPLICANT: Gray, Kevin A.
; APPLICANT: Barton, Nelson R.
; APPLICANT: Garrett, James B.
; APPLICANT: O'Donoghue, Eileen
; APPLICANT: Mathur, Eric J.
; TITLE OF INVENTION: PHYTASES, NUCLEIC ACIDS ENCODING THEM
; TITLE OF INVENTION: AND METHODS FOR MAKING AND USING THEM
; FILE REFERENCE: 09010-029007
; CURRENT APPLICATION NUMBER: US/10/156,660
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: US 09/866,379
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 09/580,515
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 09/318,528
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: US 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: US 09/259,214
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: US 08/910,798
; PRIOR FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-156-660-4

Query Match 37.2%; Score 613.5; DB 14; Length 432;
Best Local Similarity 45.3%; Pred. No. 5.7e-50;
Matches 194; Conservative 44; Mismatches 79; Indels 111; Gaps 36;

QY 2 SLILAAALAAALAP--AAAAAP-LKLEKVVILSRHGVSPTKA--LM--VSPD-WP-W 52
Db 3 AILIPFSLILPTPOSAPFAQSEPELKESVIVSRHGVRAPTKATQLMQDVTPDAPTW 62
QY 53 PV---GLTPRGAALV-LLGY-----RYFARGLL--PGCPAAG--TILADVDETR-TGOAF 99
Db 63 PVKLGWLTTPRGGLIAYLGHYQRLVADGLLAKKGCPSQGVAVIADVDERTKRTGEAF 122
QY 100 AAGLAPGCAT-VH-----AGDDPIFHGLDTG-C-LDOA---DAILAAGEGLT--ARHGL 146
Db 123 AAGLAPDCAITVHTQADTSSPDLFNLKTVGVCLDNANVTDAILSRAGGSADFTGHRQ 182
QY 147 T-----LAKVLNF--ASACLE-----CARVVGRLGFL--LASTLSEI 179
Db 183 TAFRELERVLNFPQSNLCKREKQDESCSLTQALPSKLVSDNVSLTGAVSLASMLTEI 242
```

QY 180 FLLEYAQQPMVGVGRI--SAEW-----LLHNAQF-LMNRTPYIA--RATPI--LIVTALS 228
DB 243 FLLOQAQMPGPGGRITDHSQWNTLLSLHNAQFYLLQRTPEVARSATFLLDLINAALT 302
QY 229 -----PARVVLLAGHDTNLA-LGG-LDL-WQLP-OPDETTPPGGALVPELWNR 271
DB 303 PHPPQKQAYGVTLPTSVLFIAGHDNTNLNLAGALNLTLPQOPDNTTPGGELVFERWRR 362
QY 272 -----YVRVMYQTMQDLNLEPLPRILP-----IPGC---GSAACSLSDFARLV 314
DB 363 LSDNSQWIVQSVLFTQVLTQQRDKTPLSLNTPPGVVKLTLAGCERNAQCMCSLAGTQIV 422
QY 315 -----APAC 318
DB 423 NEARIPAC 430
RESULT 13
US-09-777-566A-2
; Sequence 2, Application US/09777566A
; Patent No. US20010055788A1
; GENERAL INFORMATION:
; APPLICANT: DIVERSA CORPORATION
; APPLICANT: SHORT, Jay
; APPLICANT: KRETZ, Keith
; APPLICANT: GRAY, Kevin
; APPLICANT: BARTON, Nelson
; APPLICANT: GARRETT, James
; APPLICANT: O'DONOGHUE, Eileen
; TITLE OF INVENTION: RECOMBINANT BACTERIAL PHYTASES AND USES THEREOF
; FILE REFERENCE: DIVER1370-6
; CURRENT APPLICATION NUMBER: US/09/777,566A
; CURRENT FILING DATE: 2001-06-11
; PRIOR APPLICATION NUMBER: US 09/318,528
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: US 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: US 09/259,214
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: US 08/910,798
; PRIOR FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Escherichia coli
; NAME/KEY: misc feature
; LOCATION: (1)-(1323)
; OTHER INFORMATION: n is any nucleotide
US-09-777-566A-2

Query Match 37.1%; Score 611.5; DB 9; Length 440;
Best Local Similarity 45.1%; Pred. No. 9e-50;
Matches 193; Conservative 45; Mismatches 79; Indels 111; Gaps 36;
QY 2 SLLLAALLAALAP--AAAAEP-LKLEKVVILSRHGVRSPTKA--LM--VSPD-WP-W 52
DB 3 AILIPFLSLILPTQSAFAQSEPELKLKESVIVSRHGVRAPTKATQMLQMDVTPDAPWTW 62
QY 53 PV---GLTPRGAALV-LLGY-----RYFARGLL--PGCPAAG--TILADVDETR-TGOAF 99
DB 63 PVKLGWLTTPRGELIAYLGHYQORQLVADGLLAKKGCPSQGOVAIIADVDETRKTGEAF 122
QY 100 AAGLAPGCAI-VH-----AGDDPIFHGLDTG-C-LDOA---DAILAAGEGGLT--ARHGL 146
DB 123 AAGLAPDCAITVHTQADTSSPDPLFNPLKTCVCOLDNANVTDAILSRAGGSIAFTGHRQ 182
QY 147 T---LAKVLFNF--ASACLE-----CARVVGRIGLPL--LASTLSEI 179
DB 183 TAPRELRLVNFPSQNLCKREKQDECSLTOALPSELKVSADNVSLTGAVSLASMLTEI 242
QY 180 FLLEYAQQPMVGVGRI--SAEW-----LLHNAQF-LMNRTPYIA--RATPI--LIVTALS 228
DB 243 FLLOQAQMPGPGGRITDHSQWNTLLSLHNAQFYLLQRTPEVARSATFLLDLINAALT 302

QY 229 -----PARVVLLAGHDTNLA-LGG-LDL-WQLP-OPDETTPPGGALVPELWNR 271
DB 303 PHPPQKQAYGVTLPTSVLFIAGHDNTNLNLAGALNLTLPQOPDNTTPGGELVFERWRR 362
QY 272 -----YVRVMYQTMQDLNLEPLPRILP-----IPGC---GSAACSLSDFARLV 314
DB 363 LSDNSQWIVQSVLFTQVLTQQRDKTPLSLNTPPGVVKLTLAGCERNAQCMCSLAGTQIV 422
QY 315 -----APAC 318
DB 423 NEARIPAC 430
RESULT 14
US-09-866-379-2
; Sequence 2, Application US/09866379
; Patent No. US20020136754A1
; GENERAL INFORMATION:
; APPLICANT: DIVERSA CORPORATION
; APPLICANT: SHORT, Jay
; APPLICANT: KRETZ, Keith
; APPLICANT: GRAY, Kevin
; APPLICANT: BARTON, Nelson
; APPLICANT: GARRETT, James
; APPLICANT: O'DONOGHUE, Eileen
; TITLE OF INVENTION: RECOMBINANT BACTERIAL PHYTASES AND USES THEREOF
; FILE REFERENCE: DIVER1370-7
; CURRENT APPLICATION NUMBER: US/09/866,379
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 09/580,515
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 09/318,528
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: US 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: US 09/259,214
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: US 08/910,798
; PRIOR FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 2
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-866-379-2

Query Match 37.1%; Score 611.5; DB 9; Length 440;
Best Local Similarity 45.1%; Pred. No. 9e-50;
Matches 193; Conservative 45; Mismatches 79; Indels 111; Gaps 36;
QY 2 SLLLAALLAALAP--AAAAEP-LKLEKVVILSRHGVRSPTKA--LM--VSPD-WP-W 52
DB 3 AILIPFLSLILPTQSAFAQSEPELKLKESVIVSRHGVRAPTKATQMLQMDVTPDAPWTW 62
QY 53 PV---GLTPRGAALV-LLGY-----RYFARGLL--PGCPAAG--TILADVDETR-TGOAF 99
DB 63 PVKLGWLTTPRGELIAYLGHYQORQLVADGLLAKKGCPSQGOVAIIADVDETRKTGEAF 122
QY 100 AAGLAPGCAI-VH-----AGDDPIFHGLDTG-C-LDOA---DAILAAGEGGLT--ARHGL 146
DB 123 AAGLAPDCAITVHTQADTSSPDPLFNPLKTCVCOLDNANVTDAILSRAGGSIAFTGHRQ 182
QY 147 T---LAKVLFNF--ASACLE-----CARVVGRIGLPL--LASTLSEI 179
DB 183 TAPRELRLVNFPSQNLCKREKQDECSLTOALPSELKVSADNVSLTGAVSLASMLTEI 242
QY 180 FLLEYAQQPMVGVGRI--SAEW-----LLHNAQF-LMNRTPYIA--RATPI--LIVTALS 228
DB 243 FLLOQAQMPGPGGRITDHSQWNTLLSLHNAQFYLLQRTPEVARSATFLLDLINAALT 302
QY 229 -----PARVVLLAGHDTNLA-LGG-LDL-WQLP-OPDETTPPGGALVPELWNR 271

Db 303 PHPPQKAYGVLTPTSVLFIAGHTNTLANLGALBNWTLPGQPDNTPPGGELVFERWR 362
QY 272 -----YRVVYQTMQDLRNLEPLPRILP-----IPGC---GSEAAACSLSDPARLV 314
Db 363 LSDNSQWIOVSLVFQTLQWRDKTPLSLNTPGCEVVKTLAGCEERNAQGMCSLAGFTQIV 422
QY 315 -----APAC 318
Db 423 NEARIPAC 430

RESULT 15

US-10-034-985-2
; Sequence 2, Application US/10034985
; Publication No. US20030049815A1
; GENERAL INFORMATION:
; APPLICANT: Kretz, Keith
; TITLE OF INVENTION: NOVEL PHYTASE
; FILE REFERENCE: 09010/029003
; CURRENT APPLICATION NUMBER: US/10/034,985
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: US/09/580,515
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: 08/910,798
; PRIOR FILING DATE: 1997-08-13
; PRIOR APPLICATION NUMBER: 09/259,214
; PRIOR FILING DATE: 1999-03-01
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-034-985-2

Query Match 37.1%; Score 611.5; DB 14; Length 440;
Best Local Similarity 45.1%; Pred. No. 9e-50;
Matches 193; Conservative 45; Mismatches 79; Indels 111; Gaps 36;

QY 2 SLLAALLAALAP--AAAAAP-LKLEKVTLSRHGVSPTKA--LM--VSPD-WP-W 52
Db 3 AILIPFLSLIIPUTPOSAFQAQSEPELKSIVSRHGVRAPTKATQLMQDVPDAWPTW 62
QY 53 PV---GLTPGAAIV-LLGY---RYFARGLL--PGCPAAG--TILADVDERT--TGOAF 99
Db 63 FVKLGWLTFRGGELIAYLGHYQRLVADGLLAKGCPQSGQVAILADVDERTRKTGEAF 122
QY 100 AAGLAPGCAL-VH-----AGDDPIFHGLDTG-C-IDQA---DAILAAGEGGLT--ARGL 146
Db 123 AAGLAPDCATVHTQADTSPDPFLNPLKTVGVCLDNANVTDAILSRAGGSADFTGHRQ 182
QY 147 T-----LAKVLNF--ASACLE-----CARVVGRLGPL--LASTLSI 179
Db 183 TAFRELRLVNFQSNCLKREKQDESCSLTQALPSELKVSADNVSLTGAVSLASMLTEI 242
QY 180 FLLEYAQGPNEVGRI--SAEW-----LLHNAQF-LMNRTPYA--RATPI--LIVTALS 228
Db 243 FLQQAQGMPEPGMGRIITDHSQWNTLSLHNAQFYLLQRTPEVARSRATPLDLMAALT 302
QY 229 -----PARVLLAGHDTNLA-LGG-LDL-WOLP-QPDETTPGGALVPELMNR 271
Db 303 PHPPQKAYGVLTPTSVLFIAGHTNTLANLGALBNWTLPGQPDNTPPGGELVFERWR 362
QY 272 -----YRVVYQTMQDLRNLEPLPRILP-----IPGC---GSEAAACSLSDPARLV 314
Db 363 LSDNSQWIOVSLVFQTLQWRDKTPLSLNTPGCEVVKTLAGCEERNAQGMCSLAGFTQIV 422
QY 315 -----APAC 318
Db 423 NEARIPAC 430

Search completed: April 30, 2004, 12:55:05
Job time : 33.0244 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 30, 2004, 12:35:05 ; Search time 11.6341 Seconds
(without alignments)
2629.235 Million cell updates/sec

Title: US-10-021-723B-14

Perfect score: 1647

Sequence: 1 VSLTALLALALAPAAAA.....CGSEAACSLSDFARLVAPAC 318

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Pir1.*

2: Pir2.*

3: Pir3.*

4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	613.5	37.2	432	B36733	acid phosphatase (
2	606.5	36.8	434	H90770	phosphoanhydride p
3	606.5	36.8	444	D85633	hypothetical prote
4	590	35.8	441	AC0201	acid phosphatase (
5	377.5	22.9	414	E87316	periplasmic phosph
6	289.5	17.6	413	F90773	periplasmic glucos
7	289.5	17.6	413	B95636	periplasmic glucos
8	288.5	17.5	413	JV0087	Glucose-1-phosphat
9	278.5	16.9	413	AC0632	Glucose-1-phosphat
10	256	15.5	417	D25627	Glucose-1-phosphat
11	134.5	8.2	423	A33395	acid phosphatase (
12	127.5	7.7	421	S14742	acid phosphatase (
13	124.5	7.6	423	S06167	acid phosphatase (
14	117	7.1	502	B93353	xylofuranase PA
15	113.5	6.9	657	D70872	probable ctpd prot
16	107.5	6.5	426	E87385	chromate transport
17	106	6.4	972	E97421	NADH dehydrogenase
18	106	6.4	994	D2639	pH adaption potass
19	104	6.3	358	D72637	hypothetical prote
20	103.5	6.3	386	JH0610	acid phosphatase (
21	102.5	6.2	4613	T17409	polyketide synthas
22	102	6.2	356	E84199	hypothetical prote
23	101.5	6.2	532	E87343	ABC transporter, A
24	99	6.0	426	E83172	probable transport
25	98.5	6.0	411	S41945	hypothetical prote
26	98	6.0	402	S25077	monensin polyketid
27	98	6.0	409	D70863	probable fadE17 pr
28	98	6.0	419	D70863	hypothetical prote
29	96.5	5.9	438	S64682	acid phosphatase (

30 96 5.8 478 2 C72658
31 95.5 5.8 761 1 S52769
32 95 5.8 358 2 T36783
33 95 5.8 832 2 B70507
34 93.5 5.7 469 2 B83368
35 93.5 5.7 571 2 C84169
36 93.5 5.7 693 2 C87575
37 93.5 5.7 1182 2 T29097
38 92.5 5.6 893 2 E87244
39 92.5 5.6 1398 2 T18350
40 92 5.6 288 2 A11171
41 92 5.6 560 2 G87613
42 92 5.6 1098 2 G70697
43 92 5.6 1165 2 D59433
44 91.5 5.6 245 2 E87273
45 91.5 5.6 520 2 T36454

ALIGNMENTS

RESULT 1

B36733
acid phosphatase (EC 3.1.3.2) precursor - Escherichia coli (strain K-12)
N;Alternate names: phosphoanhydride phosphohydrolase, periplasmic; phytase 6
C;Species: Escherichia coli
C;Date: 19-Apr-1991 #sequence revision 19-Apr-1991 #text change 01-Mar-2002
C;Accession: B36733; S18018; B64839; A26534; S17960; S33278
R;Dassa, J.; Marck, C.; Boquet, P.L.
J;Bacteriol. 172, 5497-5500, 1990
A;Title: The complete nucleotide sequence of the Escherichia coli gene appA reveals s
A;Reference number: A36733; MUID:90368616; PMID:2168385
A;Accession: B36733
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-432 <DAS>
A;Cross-references: GB:M58708; NID:g145283; PIDN:AAA72086.1; PID:g145285
R;Greiner, R.; Jany, K.D.
Biol. Chem. Hoppe-Seyler 372, 664-665, 1991
A;Title: Characterization of a phytase from Escherichia coli.
A;Reference number: S18018
A;Accession: S18018
A;Molecule type: protein
A;Residues: 23-33 <GRE>
R;Blatner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: B64839
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-432 <BLAT>
A;Cross-references: GB:AF000200; GB:U00096; NID:92367111; PIDN:AAC74065.1; PID:g178721
A;Experimental source: strain K-12, substrain MG1655
R;Touati, E.; Danchin, A.
Biochimie 69, 215-221, 1987
A;Title: The structure of the promoter and amino terminal region of the pH 2.5 acid ph
A;Reference number: A26534; MUID:87271766; PMID:3038201
A;Accession: A26534
A;Molecule type: DNA
A;Residues: 1-50, 'NAGCHPRMANLAG', 65, 'T', 67-74, 'DV', 77-111, 'S', <TOU>
A;Cross-references: GB:X05471; NID:g40925; PIDN:CAA29031.1; PID:g40927
R;Dassa, J.; Fathi, H.; Marck, C.; Dion, M.; Kieffer-Bontemps, M.; Boquet, P.L.
Mol. Gen. Genet. 229, 341-352, 1991
A;Title: A new oxygen-regulated operon in Escherichia coli comprises the genes for a p
A;Reference number: S17958; MUID:92049231; PMID:1658595
A;Accession: S17960
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-17 <DA2>
A;Cross-references: GB:S63811; NID:g238656; PIDN:AAB20286.1; PID:g238659
R;Greiner, R.; Konietzny, U.; Jany, K.D.

Arch. Biochem. Biophys. 303, 107-113, 1993
A;Title: Purification and characterization of two phytases from *Escherichia coli*.
A;Reference number: S33278; MUID:93256556; PMID:8387749
A;Accession: S33278
A;Molecule type: protein
A;Residues: 23-31; A, 33-34 <GR2>
C;Comment: In addition to cAMP-mediated control, this enzyme is induced when bacterial
necrotic conditions.
C;Genetics:

A;Gene: appA
A;Map position: 25 min
A;Keywords: monomer; periplasmic space; phosphatidyl; phosphoprotein; phosphoric mon
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-112/Product: acid phosphatase #status predicted <MAT>
F;38/Active site: Arg #status predicted
F;39/Active site: His (phosphohistidine intermediate) #status predicted

Query Match 37.2%; Score 613.5; DB 2; Length 432;
Best Local Similarity 45.3%; Pred. No. 3.6e-44;
Matches 194; Conservative 44; Mismatches 79; Indels 111; Gaps 36;

QY 2 SLLAALLAALAP-AAAAAP-LKLEKVVLSHGVRSPYKA--LM--VSPD-WP-W 52
DB 3 AILIPFSLIPLTPQSAFAQSPPELKESVIVSRHGVRAPTKATQLMQDVTDPDAMP 62
QY 53 PV---GLTPRGAALV-LLGY---RYFARGLL--PGCPAAG--TILADVDETR-TGOAF 99
DB 63 PVKLGWLTFRGGELIAYLGHYQORLVADGLLTKKGCPOGOVAILADVDETRTKTGE 122
QY 100 AAGLAPGCAI-VH-----AGDDPIFGHLDTG-C-LDQA---DAILAAGEGLT--ARH 146
DB 123 AAGLAPDCAITVHTQADTSSPDPLFNPLKTVGCVQCLDNANVTDAILSRAGGSADFTGH 182
QY 147 T-----LAKVLFN--ASACLE-----CARVVGRLGFL--LASTLS 179
DB 183 TAPRELFVLPFQSNLCINREKQDECSLTQALPSELKVSADNVSLTGAVSLASMLT 242
QY 180 FLLEYAQQPMVEGWGRI--SAEW-----LLHNAQF-LMNRTPYIA--RATPI--LIVTA 228
DB 243 FLIQAQGMPEPGWGRITDSHQWNTLLSLHNAQFYLLQRTPEVARSRAATPLLDLITMIA 302
QY 229 -----PARVLLAGHDNTLA-LGG-LDL-WOLP-OPDETTPGGALVVELNW 271
DB 303 PHPPKQAYGVTLPTSVLFIAGHDNTLANLGALNLWTLPGQPDNTPPGSELVFERW 362
QY 272 -----YRVVMYQMDQLRNLPLPRILP-----IPGC---GSEAAACSLSDPAR 314
DB 363 LSDNSQWIVQSLVFTLQMDKTPSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQIV 422
QY 315 -----APAC 318
DB 423 NEARIPAC 430

RESULT 2
H90770
Phosphoanhydride phosphorylase [imported] - *Escherichia coli* (strain O157:H7, substrain
C;Species: *Escherichia coli*
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: H90770
R;Hayashi, T.; Makino, K.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gene
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: H90770
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-434 <HAY>
A;Cross-references: GB:BA000007; PIDN:BA034559.1; PID:G13360596; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RMD 0509952
C;Genetics:
A;Gene: Ecs1136

Query Match 36.8%; Score 606.5; DB 2; Length 434;
Best Local Similarity 44.9%; Pred. No. 1.4e-43;
Matches 193; Conservative 43; Mismatches 81; Indels 113; Gaps 36;

QY 2 SLLAALLAALAPAAAAA---EP-LKLEKVVLSHGVRSPYKA--LM--VSPD-WP 51
DB 3 AILIPFSLIPLTPQSAFAQSPPELKESVIVSRHGVRAPTKATQLMQDVTDPDAMP 62
QY 52 -WPV---GLTPRGAALV-LLGY---RYFARGLL--PGCPAAG--TILADVDETR-TGO 97
DB 63 NWPVKLGWLTFRGGELIAYLGHYQORLVADGLLTKKGCPOGOVAILADVDETRTKTGE 122
QY 98 AFAAGLAPGCAI-VH-----AGDDPIFGHLDTG-C-LDQA---DAILAAGEGLT--ARH 144
DB 123 AFAAGLAPDCAITVHTQADTSSPDPLFNPLKTVGCVQCLDNANVTDAILSRAGGSADFTGH 182
QY 145 GLT-----LAKVLFN--ASACLE-----CARVVGRLGFL--LASTLS 177
DB 183 RQAFRELFVLPFQSNLCINREKQDECSLTQALPSELKVSADNVSLTGAVSLASMLT 242
QY 178 EIFLLEYAQQPMVEGWGRI--SAEW-----LLHNAQF-LMNRTPYIA--RATPI--LIVTA 226
DB 243 EIFLLOQAQGMPEPGWGRITDSHQWNTLLSLHNAQFYLLQRTPEVARSRAATPLLDLITMIA 302
QY 227 LS-----PARVLLAGHDNTLA-LGG-LDL-WOLP-OPDETTPGGALVVELW 269
DB 303 LTPHPKQAYGVTLPTSVLFIAGHDNTLANLGALNLWTLPGQPDNTPPGSELVFERW 362
QY 270 NR-----YRVVMYQMDQLRNLPLPRILP-----IPGC---GSEAAACSLSDPAR 312
DB 363 RRLSDNSQWIVQSLVFTLQMDKTPSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQ 422
QY 313 LV-----APAC 318
DB 423 IVNEARIPAC 432

RESULT 3

D85633
hypothetical protein appA [imported] - *Escherichia coli* (strain O157:H7, substrain EDI
C;Species: *Escherichia coli*
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: D85633
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayh
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodac
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: D85633
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-444 <STO>
A;Cross-references: GB:AE005174; NID:G12514245; PIDN:AAG55528.1; GSPDB:GN00145; UWGP:2
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: appA

Query Match 36.8%; Score 606.5; DB 2; Length 444;
Best Local Similarity 44.9%; Pred. No. 1.4e-43;
Matches 193; Conservative 43; Mismatches 81; Indels 113; Gaps 36;

QY 2 SLLAALLAALAPAAAAA---EP-LKLEKVVLSHGVRSPYKA--LM--VSPD-WP 51
DB 13 AILIPFSLIPLTPQSAFAQSPPELKESVIVSRHGVRAPTKATQLMQDVTDPDAMP 72
QY 52 -WPV---GLTPRGAALV-LLGY---RYFARGLL--PGCPAAG--TILADVDETR-TGO 97
DB 73 NWPVKLGWLTFRGGELIAYLGHYQORLVADGLLTKKGCPOGOVAILADVDETRTKTGE 132
QY 98 AFAAGLAPGCAI-VH-----AGDDPIFGHLDTG-C-LDQA---DAILAAGEGLT--ARH 144
DB 133 AFAAGLAPDCAITVHTQADTSSPDPLFNPLKTVGCVQCLDNANVTDAILSRAGGSADFTGH 192

QY 145 GLT---LAKVNF---ASACLE-----CARVVGRLGPL-LASTUS 177
Db 193 RQAFRELERVNFPOSNLCNREKQDECSLTQALPSELKVSADNVSLTGAVSLASMLT 252
QY 178 EIFLLYAOQPMVGVGRI--SAEW-----LLHNAQF-LMNRTPYIA--RATPI--LIIVTA 226
Db 253 EIFLLOQAQMPGPGWGRITDSHQWNTLLSLHNAQFYLLQRTPEVARSATPLDLDIMTA 312
QY 227 LS-----PARVVLLAGHDNTLA-LGG-LDL-WQLP-QDETTPPGGALVFELW 269
Db 313 LTHPPQKQAYGVTLPFVSFLFIAGHDNTLANLGALELNLWLPQPDNTPPGGELVFERW 372
QY 270 NR-----YVRVMTQMDQLNRLEPLPRILP-----IPGC-----GSEACSLSDPAR 312
Db 373 RRLSDNSQWISLVFQTLQMRDKTPLSLNTPPGEVYKTLTACBERRNAQGMCSLAGFTQ 432
QY 313 LV-----APAC 318
Db 433 IVNEARIPAC 442

RESULT 4
AC0201
acid phosphatase (EC 3.1.3.2) [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001
C:Accession: AC0201
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
Ill, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,
Nature 413, 523-527, 2001
A>Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AC0201
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-441 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC90470.1; PID:gl5979685; GSPDB:GN00175
A;Gene: YPoi648
C;Keywords: phosphoric monoester hydrolase

Query Match 35.8%; Score 590; DB 2; Length 441;
Best Local Similarity 43.4%; Pred. No. 3.5e-42;
Matches 187; Conservative 42; Mismatches 88; Indels 114; Gaps 34;

QY 1 VSLLLAALLAAL-APAAAAEPLKLEKVVLSRHGVRSPYK--ALM--VSPD-WP-WP 53
Db 10 LSGVLMLSLGLAATAPVAAEPGSGYTLERVVLSRHGVRSPYKQTQLMNDVTPDRWQWP 69
QY 54 VG---LTPRGAALVLL-----GYRYFARGLL-PCGPAAGTIL--ADVDERTR-TGQAFAA 101
Db 70 VKAGYLTFRGAELVTLMGGFGVDYFRSLGILLAGCPAEGGVYAQADIDQTRLTGQAFLD 129
QY 102 GLAPGCAI-VH-----AGDDPIFHGLDTG-C-LD--QADAILAAGEGG-----LTARHGLT 147
Db 130 GVAPGCGLTVHQADLKKTDPLFHFVAGVCKLDAADKAIBEQLGGLDTVTSQRYAKP 189
QY 148 LAK---VLNFASA--CLECAR-----VVRGLGPL-LASTLSEI 179
Db 190 FAQMGDVNFPAASYCKSLQOQGTCDFAHPAANEVNVNKEGTVKVTLSGLPLASSTLGEI 249
QY 180 FLLEYAQPMVGVGRI--SAEWL-----LHNAQF-LMNRTPYIAR--ATPIL--IVTAL- 227
Db 250 FLLQNAQAMPEVAMQRLKGAENWVSLSLHNAQFNLMKTPYIARHKGTPLLQOQIDTALT 309
QY 228 -----SPAVVLLAGHDNTLA-----LGGDLWLQLP-QDETTPPGGALVFELW 269
Db 310 LQDQAQGKLPISAGNRVLFVGGHDNTIANIAGLMGAD-WQLPEQPDNTPPGGELVFELW 368
QY 270 -----NRY--VRVMTQMDQLNRLEPLPRILP-----IPGC-----GSEACSLSDP- 310

Db 369 QNPNDHQRYVAVKMFYQTMQDLNRNAEKLDLKNNPAGIISVAVAGCENNGDDKLCLELDTFQ 428
QY 311 ---ARLVAPAC 318
Db 429 KKVAVKVIETPAC 439

RESULT 5
E87316
periplasmic phosphoanhydride phosphohydrolase [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: E87316
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, C.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolc
N. J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.N
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A>Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: E87316
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-414 <STO>
A;Cross-references: GB:AB005673; NID:gl3421731; PIDN:AAK22529.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC0542

Query Match 22.9%; Score 377.5; DB 2; Length 414;
Best Local Similarity 35.9%; Pred. No. 2.9e-24;
Matches 148; Conservative 47; Mismatches 112; Indels 105; Gaps 29;

QY 4 LLAALLAALLAALAPAAAAEPLKLEKVVLSRHGVRSPYKALMVSPD-----WP--- 51
Db 5 LRIAVVTLALASAGASAE--TLEKVVLSRHGVR-----AMSSPERLEEASARWPPE 58
QY 52 WPMV-LTPRGAALV-LG--YR--YFARGLL-PCGPAAGTILADVDET-RTGQAFAGL 103
Db 59 VPAGHLTARGETLVARMGDYRRHYAAQGLLKPGDCASVYAWANVTQRTIATAKAYRTL 118
QY 104 APGCAI-----VHAGD-DPIFHGLDTGCLDQADAILA-----AGE--GGLTARHGLTAKVL 152
Db 119 APGCPVTNTVGEINIDFEPFVAGIV-KADHALARAANVAGRGCDLTANSASHNQAE 177
QY 153 NFASACLECAR-----VVRGLQF-LMNRTPYIAR-----ATPILIVTALS 228
Db 178 QLDALLMQCDKPCPPAPGKRRVFDKPGFVDGEBELAGLSPGPAFASGVTSLLMAWADG 237
QY 188 P--MEVGHGRISAE-----WLLHNAQF-LMNRTPYIAR-----ATPILIVTALS 228
Db 238 RDFAGLGRKSLDEEALTRSFLLHQAEFDLRLTPVARTLHAGLADRLAATLRDGAAG 297
QY 229 P--ARVVLLAGHDNTLA-LGGL--DLWQLP--QDETTPPGGALVFELW-----RYVRM 276
Db 298 FVDARLVIIAGHDGTLASLGGLLRMEWTLPGVQPNQIQPGGALVFERWRDDGVVRVR 357
QY 277 Y--QTMQDLNRLEPLPRILP-----IPGCSENA-----CSLSDPARLVAPA 317
Db 358 FTGQSLQLNRNMTALDAKTPPLSAPFVQCGGTATPAFDCRLDEPFTVVRGA 409

RESULT 6
F90773
periplasmic glucose-1-phosphatase [imported] - Escherichia coli (strain O157:H7, subser
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: F90773
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.C
gasaawa, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and ser
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: F90773
A;Status: preliminary

A:Molecule type: DNA
A:Residues: 1-413 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA34581.1; PID:g13360618; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIND 050952
C:Genetics:
A:Gene: ECs1158

Query Match 17.6%; Score 289.5; DB 2; Length 413;
Best Local Similarity 28.6%; Pred. No. 8e-17;
Matches 116; Conservative 71; Mismatches 116; Indels 103; Gaps 29;

QY 1 VSLLLAALLAALAPAAAAPLEKLVILSHGVRSPT-----KALMVS-----PDWP 51
DB 6 IAAAVAGVLLASNAQAQTVEGYQLQVLMMSRNLRLAPLANNVSVLEQSTPNKWPED 65
QY 52 WPGV-LTPRGAAL-VLLGY---RYFA-RGLLPG---CPAAGTILADVDERTR---TGQAF 100
DB 66 VPGQLTTKGGVLEVMGHYRWELAEQGMVKSCEPPPTVYVANSQRTVATAQFFI 125
QY 101 AGLAPCAIVHAGDDPIFHGLDTGCLDQ-----ADAILAAGEGGLTARHGLT 147
DB 126 TGAFPGC-----DIPVHQQEKMGTMPTFNPVITDSDAFAFSEQAAMEKEL-SKLQUT 178
QY 148 -----LAKVLNF--ASACLE---CARVGR-----LGPL-LASTLSIFLLE 183
DB 179 DSYQLLEKIVNYKDSACKKQCSLVGKNTFSKAKYQCPGVSGPLKVGNSLVDAFTLQ 238
QY 184 YAGQ-PM-EVWGRISA--EW-----LLHNAQFLMRTPIYAR--ATPI-----LIVT 225
DB 239 YIEGFPMQVANGKEIKSDQKQVLSKLKNGYQDSLFTSEVARNVAKPLVSYIDKALVTD 298
QY 226 ALSPARVLLAGHDTNLA--LGGDL--WQL-PQDPTPPGALVPFLW-----NR---Y 272
DB 299 RTSAPKITVLVGHDSNIASLLTALDFKPYQLHDQNERITIGGKIVFQRWRDSKANEDLMK 358
QY 273 VRVYMTDQLNLEPLPR-----ILPIPGC--GSEAAACSLSDP 310
DB 359 IEVYQSAEQLRNADALTLOAPARVLTLELSCPIDADGFCPMDKF 404

RESULT 7

B85636
periplasmic glucose-1-phosphatase [imported] - Escherichia coli (strain O157:H7, substrain
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: B85636
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 402, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11208551
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-413 <STO>
A:Cross-references: GB:AE005174; NID:g12514273; PIDN:AAG55550.1; GSPDB:GN00145; UWGP:214
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: agp

Query Match 17.6%; Score 289.5; DB 2; Length 413;
Best Local Similarity 28.6%; Pred. No. 8e-17;
Matches 116; Conservative 71; Mismatches 116; Indels 103; Gaps 29;

QY 1 VSLLLAALLAALAPAAAAPLEKLVILSHGVRSPT-----KALMVS-----PDWP 51
DB 6 IAAAVAGVLLASNAQAQTVEGYQLQVLMMSRNLRLAPLANNVSVLEQSTPNKWPED 65
QY 52 WPGV-LTPRGAAL-VLLGY---RYFA-RGLLPG---CPAAGTILADVDERTR---TGQAF 100
DB 66 VPGQLTTKGGVLEVMGHYRWELAEQGMVKSCEPPPTVYVANSQRTVATAQFFI 125
QY 101 AGLAPCAIVHAGDDPIFHGLDTGCLDQ-----ADAILAAGEGGLTARHGLT 147

DB 126 TGAFPGC-----DIPVHQQEKMGTMPTFNPVITDSDAFAFSEQAAMEKEL-SKLQUT 178
QY 148 -----LAKVLNF--ASACLE---CARVGR-----LGPL-LASTLSIFLLE 183
DB 179 DSYQLLEKIVNYKDSACKKQCSLVGKNTFSKAKYQCPGVSGPLKVGNSLVDAFTLQ 238
QY 184 YAGQ-PM-EVWGRISA--EW-----LLHNAQFLMRTPIYAR--ATPI-----LIVT 225
DB 239 YIEGFPMQVANGKEIKSDQKQVLSKLKNGYQDSLFTSEVARNVAKPLVSYIDKALVTD 298
QY 226 ALSPARVLLAGHDTNLA--LGGDL--WQL-PQDPTPPGALVPFLW-----NR---Y 272
DB 299 RTSAPKITVLVGHDSNIASLLTALDFKPYQLHDQNERITIGGKIVFQRWRDSKANEDLMK 358
QY 273 VRVYMTDQLNLEPLPR-----ILPIPGC--GSEAAACSLSDP 310
DB 359 IEVYQSAEQLRNADALTLOAPARVLTLELSCPIDADGFCPMDKF 404

RESULT 8

JV0087
glucose-1-phosphatase (EC 3.1.3.10) precursor - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 01-Mar-2002
C:Accession: JV0087; H64841
R:Pradel, E.; Marck, C.; Boquet, P.L.
J. Bacteriol. 172, 802-807, 1990
A:Title: Nucleotide sequence and transcriptional analysis of the Escherichia coli agp
A:Reference number: JV0087; MUID:90130318; PMID:2153860
A:Accession: JV0087
A:Molecule type: DNA
A:Residues: 1-413 <PRA>
A:Cross-references: GB:M33807; NID:g145217; PIDN:AAA23426.1; PID:g145218
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: H64841
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-413 <BLAT>
A:Cross-references: GB:AE000202; GB:U00096; NID:g1787233; PIDN:AAC74087.1; PID:g178723
A:Experimental source: strain K-12, substrain MGL655
C:Genetics:
A:Gene: agp
A:Map position: 23 min
C:Function:
A:Description: essential for growth in a high-phosphate medium containing glucose-1-ph
A:Note: Optimal at low pH
C:Keywords: homodimer; periplasmic space; phosphohistidine; phosphoprotein; phosphoric
F:1-22/Domain: signal sequence #status predicted <sig>
F:23-413/Product: glucose-1-phosphatase #status predicted <MAT>
F:40/Active site: His (phosphohistidine intermediate) #status predicted
F:311/Active site: His #status predicted

Query Match 17.5%; Score 288.5; DB 2; Length 413;
Best Local Similarity 28.6%; Pred. No. 9.7e-17;
Matches 116; Conservative 71; Mismatches 116; Indels 103; Gaps 29;

QY 1 VSLLLAALLAALAPAAAAPLEKLVILSHGVRSPT-----KALMVS-----PDWP 51
DB 6 IAAAVAGVLLASNAQAQTVEGYQLQVLMMSRNLRLAPLANNVSVLEQSTPNKWPED 65
QY 52 WPGV-LTPRGAAL-VLLGY---RYFA-RGLLPG---CPAAGTILADVDERTR---TGQAF 100
DB 66 VPGQLTTKGGVLEVMGHYRWELAEQGMVKSCEPPPTVYVANSQRTVATAQFFI 125
QY 101 AGLAPCAIVHAGDDPIFHGLDTGCLDQ-----ADAILAAGEGGLTARHGLT 147
DB 126 TGAFPGC-----DIPVHQQEKMGTMPTFNPVITDSDAFAFSEQAAMEKEL-SKLQUT 178

QY 148 -----LAKVLNF--ASACLE---CARVVGK-----LGPI-LASTLSIFLLE 183
DB 179 DSYQLLEKIVNYKDSACKKQCSLVGDKNTFSKAYQOBPGVSGPLKVGNSLVDAFTLQ 238
QY 184 YAOQ-PM-EVWGGRISA--EW-----LLHNAQFLMNRTPYIAR--ATPI-----LIVT 225
DB 239 YIEGFPMDQVANGIKSDQKVKSLKNGVQDSLFTSPEVARNVAKPLVYIDKALVTD 298
QY 226 ALSPARVVLVAGHTNLA--LGGLDL--WQL-POPDTPGGALVFLW-----NR--Y 272
DB 299 RTSAPKITVLVGHDSNIALTLALDFKPYQLHDQNTPTGGKIVFORMHDSKANRDLMK 358
QY 273 VRVMYQTMQDLNLEPLPR-----ILPIPGC--GSEAAACSLSDP 310
DB 359 IEYVQSAEQLRNADALTLOPAQRVILESGCPIDADGFCPMDKF 404

RESULT 9
AG0632
C:Species: Salmonella enterica subsp. enterica serovar Typhi
C:Note: This species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AG0632
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher, T.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AG0632
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-413 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD08242.1; PID:g16502289; GSPDB:GN00176
C:Genetics:
A:Gene: STY1153

Query Match 16.9%; Score 278.5; DB 2; Length 413;
Best Local Similarity 30.6%; Pred. No. 6.8e-16;
Matches 125; Conservative 59; Mismatches 128; Indels 97; Gaps 29;

QY 4 LLALLAALAPAAAAA-----EPKLEKVVILSRHGVRSPT-----KALMVS-----PDW 50
DB 5 LLAVAVAGAVLSSAVOQTTPGYQLQVLMMSRHLNRAFLNNGNVLAQSTPNAPFAW 64
QY 51 FWPVG-LTPRGAAL-VLLGYR---YFARGLLPG--CPAAGTILADVDERTR---TQAF 99
DB 65 DVPGGQLTKGVLEVYMGHTREWLVAQGLIPSGCEPAPDTVYVANSQRTVATQFF 124
QY 100 AAGLAPGCAI-VHAGD-----DPIFHGLDTGCLDQADAILAAGEGGLTARHGL----- 146
DB 125 ITSAPFGCDIPVHHQEKMGTMPTFNPTVD--DAAFRQAVQAMEKARSQHLDSYK 182
QY 147 TLAKVLNF--ASACLE---CARV-----VRLGLPL--LASTLSIFLLEVAQ 187
DB 193 LLEQHTIQDSFCKEHCQCLSDAKOTFSANYQOBPGVQPLKVGNSLVDAFTLQYEG 242
QY 198 -PM-EVWGGRISA--SAEW-----LLHNAQFLMNRTPYIAR--ATP-----LIVTALSP 229
DB 243 FPMQDQVANGIHTDRQKVKSLKNGYQDSLFTSFTVARNVAAPLVKYIDKVLVAERVA 302
QY 230 ARVVLVAGHTNLA--LGGLDL--WQL-POPDTPGGALVFLW-----NR-----YVRVM 276
DB 303 PKVTVLVGHDSNIALTLALDFKPYQLHDQVETPTGGQLVFORWHQGNANRDLMKIEYV 362

277 YQTMQDLNLEPLPR-----ILPIPGCSEA--ACSLSDFAIRLVA 317
363 YQSAQLNAEALTLKSPAQRVLELKGCPVDANGFCPLDFDNVMTA 411

RESULT 10

S25627
glucose-1-phosphatase (EC 3.1.3.10) precursor - Providencia rettgeri
C:Species: Providencia rettgeri
C:Date: 25-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 08-Oct-1999
C:Accession: S25627
R:Riccio, M.L.; Chiesurin, A.; Lombardi, G.; Satta, G.
submitted to the EMBL Data Library, September 1992
A:Reference number: S25627
A:Accession: S25627
A:Molecule type: DNA
A:Residues: 1-417 <RIC>
A:Cross-references: EMBL:X68201; NID:g45771; PIDN:CAA48288.1; PID:g45772
A:Experimental source: strain PV7
C:Genetics:
A:Gene: ggp
C:Keywords: periplasmic space; phosphohistidine; phosphoprotein; phosphoric monoester 1
F;1-29/DNA: signal sequence #status predicted <SIG>
F;30-417/Product: glucose-1-phosphatase #status predicted <MAT>
F;42/Active site: His (phosphohistidine intermediate) #status predicted
F;314/Active site: His #status predicted

Query Match 15.5%; Score 256; DB 2; Length 417;
Best Local Similarity 29.0%; Pred. No. 5.5e-14;
Matches 119; Conservative 57; Mismatches 140; Indels 94; Gaps 27;

QY 1 VSLLAALLAALAPAAAAA-----PLKLEKVVILSRHGVRSPT-----TKALMVS-----PD 49
DB 6 LTLCLSAAL-FAPIAPTMASTDNQADWLDQVLSRHLNRAPIVNTGILTEVTDKWPFD 64
QY 50 WWPVG-LTPRGAALVLLGYRFAAGLLPG-----CPAAGT---ILADVDERTRTQ 97
DB 65 WDAKSGYLTQGGALEVYMGHTREWLVAQGLIPSGCEPAPDTVYVANSQRTVATQ 124
QY 98 AFAAGLAPGCAI-VH-----AGDDPIFHGLDT-GCLDQADAILAAGEG---GLTARHGL- 146
DB 125 FPAAGAPGCKVNIHQEIGKMDVFNPIIINGSPFKQKALAAWDDVLLKSLKAGYE 184
QY 147 TLAKVLNF-----SACLECARVVGELGL--LASTLSIFLLEVAQ 187
DB 185 ELDTVLNIKDSQCKTKDLCLNDSQKNSFIIEADKPGVSGPLKANSADDAIDQYEG 244
QY 188 -PMEVWGGRISA--EW-----LLHNAQFLMNRTPYIAR--ATPI-----LIVTALS 228
DB 245 FPAQDQVANGLVDTPEKVKLNTLKAYQETLFTPKIAKNVAHPILNIDKGFVSDKGE 304
QY 229 PARVVLVAGHTNLA--LGGLDL--WQLPQP--DETTPGGALVFLW-----RYRV 275
DB 305 TAKFPLVGHDSNIALSAMSANDFKPYLAQVETPTGGKLVFORWTDKQKDFMKVEY 364
QY 276 YQTMQDLR-----NLE-PLLPRIPLPGC--GSEAAACSLSDFAIRLVA 317
DB 365 YQTMQDLRDNAYLSLETTPKRVTLKDCPVDKNGYCSWEDFKVMKA 414

RESULT 11
A33395
acid phosphatase (EC 3.1.3.2) precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 29-Jan-1990 #sequence_revision 29-Jan-1990 #text_change 22-Jun-1999
C:Accession: A33395
R:Himeno, M.; Fujita, H.; Noguchi, Y.; Kono, A.; Kato, K.
Biochem. Biophys. Res. Commun. 162, 1044-1053, 1989
A:Title: Isolation and sequencing of a cDNA clone encoding acid phosphatase in rat liv
A:Reference number: A33395; MUID:89350910; PMID:2764916
A:Accession: A33395
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-423 <HIM>
A:Cross-references: GB:M27893; NID:G202933; PIDN:AAA40744.1; PID:G202934
C:Superfamily: mammalian acid phosphatase
C:Keywords: phosphohistidine; phosphoprotein; phosphoric monoester hydrolase
F;41/Active site: Arg #status predicted
F;42/Active site: His (phosphohistidine intermediate) #status predicted

Query Match 8.2%; Score 134.5; DB 2; Length 423;
 Best Local Similarity 25.3%; Pred. No. 0.001;
 Matches 95; Conservative 33; Mismatches 157; Indels 91; Gaps 20;

QY 4 LLAALLAALAPAAAAEPLKLVILSRHGVSPTKALMVSP--DWPVVG---LTP 58
 DB 13 LLQFLLMCLMVPPIQARSRAF--VTLRYHGRDSRVKPKDYQBEKWPQGGQUTK 70
 QY 59 RGAAL-----VLGVRYPARGLLPGCPAAGTTL--ADVDERTRTGOAFAAGLAPGCAIV 110
 DB 71 EGMQLHWELGQALRORY--HGFNLASYHQRYVYVSTDFDRTLSAEANLAGLPPPTVEQ 128
 QY 111 H-----AGDDPIFHGLDTG-----CLDQADAI-L 133
 DB 129 HFNPNISWQIPVHTVPTEDRLKFLPGLCPCEPRYEQQLQNETROTPEYQNMISQNAQFLDM 188
 QY 134 RAGEGLTARHGLTTLAKVNFASACLECARVVRGLGPILLAS-----TLSEI---FLLEY 184
 DB 189 VANETGLM--NLLETINWVDT-LFCEQTHGLLLPPWASPTQVQALSQKDSFSLFLP 244
 QY 185 AQGPMEVWGGRISAEWLLHNAQFLMNRTPYIARATPILIVTALS PARVLLAGHDTNLAL 244
 DB 245 GIHD-QVQKARLQGGVLL--AQILKNLT-----LMATTSQPPKLLVYSAHDTTLVA 292
 QY 245 GGLDLWLPQDETTPGGALVFLW-----NRYVRYVYQTMQDLRLNLEPLPRILPIPGCG 300
 DB 293 LQWAL-NVYNGKQAPYASCHIFELYQEDNGNFSVEMY-----FRNDSKAPWPLTLPGCP 346
 QY 301 SEACSLSDFAFLVAP 316
 DB 347 HR--CPLQDFLRLTEP 360

RESULT 12
 S14742
 acid phosphatase (EC 3.1.3.2) precursor - mouse (fragment)
 C:Species: Mus musculus (house mouse)
 C>Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 31-Mar-2000
 C:Accession: S14742
 R:Geier, C.; von Figura, K.; Pohlmann, R.
 Biol. Chem. Hoppe-Seyler 372, 301-304, 1991
 A:Title: Molecular cloning of the mouse lysosomal acid phosphatase.
 A:Reference number: S14742; MUID:91282986; PMID:2059337
 A:Accession: S14742
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-421 <GEI>
 A:Cross-references: EMBL:X57199; NID:G52870; PIDN:CAA40485.1; PID:G52871
 C:Superfamily: mammalian acid phosphatase
 C:Keywords: lysosome; phosphoric monoester hydrolase

Query Match 7.7%; Score 127.5; DB 2; Length 421;
 Best Local Similarity 23.5%; Pred. No. 0.004;
 Matches 92; Conservative 32; Mismatches 145; Indels 123; Gaps 19;

QY 6 AALL-----ALAAPAAAAEPLKLVILSRHGVSPTKALMVSP-----DWP----- 51
 DB 9 AALLQFLLMCLMVPPIQARSRAF--VTLRYHGRDSRVKPKDYQBEKWPQGGQUTK 64
 QY 52 -----WPVGLTRPGAALVLLGYRY-----FARGLLPGCPAAGTTLADVDERTRTG 96
 DB 65 QLTKEGMLQHWELGQALRORYHGFNTSYHQRYVYRS-----TDFDRLTMSA 112
 QY 97 QAFAGLAPGCAIVH-----AGDDPIFHGL----- 121
 DB 113 EANLAGLFPPEVNEVQHFSPNISWQIPVHTVPTEDRLKFLPGLCPCEPRYEQQLQNETROTPE 172
 QY 122 -DTGCLQADAI-LAAGEGGLTARHGLTTLAKVNFASACLECARVVRGLGPILLAS----- 174
 DB 173 YQNSIQNAQFLMNVANETGLT---NLTLETINWVDT-LFCEQTHGLLLPPWASPTQVQ 228
 QY 175 -----TLSEIFLLEYAQGPMEVWGGRISAEWLLHNAQFLMNRTPYIARATPILIVTALS 228

DB 229 RLSQLKDFSLFLFGIHE---QVQKARLQGGVLL--AQILKNLT-----LMATTSQ 274
 QY 229 PARVLLAGHDTNLALGGLDLWLPQDETTPGGALVFLW-----NRYVRYVYQTMQDLR 284
 DB 275 FKLLVYSADITLVALQWAL-NVYNGKQAPYASCHIFELYQEDNGNFSVEMY-----FR 328
 QY 285 NLSPFLPRILPIPGCGSEACSLSDFAFLVAP 316
 DB 329 NDSKAPWPLTLPGCPHR--CPLQDFLRLTEP 358

RESULT 13
 S06167
 acid phosphatase (EC 3.1.3.2) ACP2 precursor [validated] - human
 N:Alternate names: acid phosphatase, lysosomal
 C:Species: Homo sapiens (man)
 C>Date: 04-Dec-1992 #sequence_revision 01-Dec-1995 #text_change 08-Dec-2000
 C:Accession: S06167; S05525; S01155
 R:von Figura, K.
 submitted to the EMBL Data Library, June 1989
 A:Reference number: S06167
 A:Accession: S06167
 A:Molecule type: DNA
 A:Residues: 1-423 <VON>
 A:Cross-references: EMBL:X15525; NID:G34239; PIDN:CAA33542.1; PID:G1199524
 R:Geier, C.; von Figura, K.; Pohlmann, R.
 Eur. J. Biochem. 183, 611-616, 1989
 A:Title: Structure of the human lysosomal acid phosphatase gene.
 A:Reference number: S05525; MUID:89377828; PMID:2776754
 A:Accession: S05525
 A:Molecule type: DNA
 A:Residues: 1-29 <GEI>
 R:Pohlmann, R.; Krentler, C.; Schmidt, B.; Schroeder, W.; Lorkowski, G.; Culley, J.; M
 EMBL J. 7, 2343-2350, 1988
 A:Title: Human lysosomal acid phosphatase: cloning, expression and chromosomal assign
 A:Reference number: S01155; MUID:89052645; PMID:3191910
 A:Accession: S01155
 A:Molecule type: mRNA
 A:Residues: 1-423 <POH>
 A:Cross-references: EMBL:X12548; NID:G34262; PIDN:CAA31064.1; PID:G34263
 A:Note: part of this sequence, including the amino end of the mature protein, was conf
 C:Genetics:
 A:Gene: GDB:ACP2
 A:Cross-references: GDB:119963; OMIM:171650
 A:Map position: 11p11.2-11p11.11
 A:Introns: 38/3; 70/3; 99/3; 150/3; 183/3; 213/3; 259/1; 285/3; 321/2; 380/1
 C:Function:
 A:Description: catalyzes the hydrolysis of a wide range of phosphate esters
 C:Superfamily: mammalian acid phosphatase
 C:Keywords: glycoprotein; phosphonitidine; phosphoprotein; phosphoric monoester hydric
 F:1-30/Domain: signal sequence #status Predicted <SIG>
 F:31-423/Product: acid phosphatase ACP2 #status experimental <MAT>
 F:41/Active site: Arg #status Predicted
 F:42/Active site: His (phosphohistidine intermediate) #status predicted
 F:92,133,167,177,191,267,322,331/Binding site: carboxylate (Asn) (covalent) #status f
 F:159-370,212-310,345-349/Disulfide bonds: #status predicted

Query Match 7.6%; Score 124.5; DB 1; Length 423;
 Best Local Similarity 23.9%; Pred. No. 0.0071;
 Matches 95; Conservative 32; Mismatches 136; Indels 135; Gaps 23;

QY 4 LLAALLA--LAALAPAAAAEPLKLVILSRHGVSPTKALMVSP-----DWP----- 51
 DB 13 LLQALLGVNLVMPPTARS---LRFVTLRYHGRDSRVKPKDYQBEKWPQGGQUTK 68
 QY 52 -----WPVGLTRPGAALVLLGYRY-----FARGLLPGCPAAGTTLADVDERTRTGOA 98
 DB 69 TKEGMLQHWELGQALRORYHGFNTSYHQRYVYRS-----TDFDRLTMSA 116
 QY 99 FAAGLAPGCAI-----VHA-----GDDPIFHGL----- 121
 DB 117 NLAGLFPPEVNEVQHFSPNISWQIPVHTVPTEDRLKFLPGLCPCEPRYEQQLQNETROTPEYQ 176

QY 122 -----DTGCLDQADAILAAGEGGLTARHGLTAKVNFASACLECARVVG-RLGPLLA-S 174
Db 177 NERRSNAQFLD-----MWANETGLT---DLLETWVNVYDT-LFCEQTHGLRLPWPASQ 227
QY 175 TLSEIFLLE-----YAQGMVEVGMGRISAEMLLHNAQFLMNRTPYIARATPILIV 224
Db 228 TMQRLSRLKDFSPRFLFGIYQAQEK-----ARLQGGVLL--AQTRKULT-----LMA 272
QY 225 TALSPARVLLAGHDNTNAL--CGLDLWQLPQDPTPGGALVFELW-----NRYVRVMYQ 278
Db 273 TTSQLEKLLVYSHDITLVALQWALDVI---NCEQAPYASCHIFELYQEDSGNFSVEMY- 328
QY 279 TMDQLRNLPLPRILPIPGCGSEAAACSLSDFARLVAP 316
Db 329 ----FRNESDKAPWPLSLPGCPHR--CPLQDFLRLTEP 360

RESULT 14
E83353
xylulose kinase PA2343 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Nov-2001
R:Accession: E83353
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Boman, S.; Yuan, Y.; Brody, L.D.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lam, J.; Lory, S.; Olson, M.V.
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
A:Reference number: A82950; MUID:2043737; PMID:10984043
A:Accession: E83353
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-502 <STO>
A:Cross-references: GB:AE004660; GB:AE004091; NID:99948372; PIDN:AAG05731.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: mtly; PA2343
C:Superfamily: xylulokinase

Query Match 7.1%; Score 117; DB 2; Length 502;
Best Local Similarity 27.0%; Pred. No. 0.037;
Matches 77; Conservative 24; Mismatches 118; Indels 66; Gaps 12;
QY 67 QY-----RYFARGLLPGCPAAGTILADVDRTRTG-----QAFAAGLAPGCAIVHA 112
Db 190 GFVDRRTWASDVLELVFPGGRLAALPELIEFCAGICGNLRPEAAALGLAETHRVAGG 249
QY 113 GDDPIFHGLDTG-----CLDQADAILAAGEGGLTARHG-----LTLAKV 151
Db 250 GGDNMALAIAGTGNTRPGLLTASLTSGTSLAYASRPLVSPHGEATFCASSGGWLPACT 309
QY 152 LNFASACLECARVVGRLGFLPILASTLSIFLLEYAOGPMVEVGMGRISAEMLLHNAQFLMNR 211
Db 310 NMLTGAC-----GLVQDLHLHLDDEFSRLA-AQAP-----VGAEGLLMLPFFDGER 354
QY 212 TPYIARATPILIT-VTALSFAVRLVLAGHDNTNALG---GLDLWQ---LPQDETPPGGAL 264
Db 355 VPALPHASASLHGMTAANLSRANLCRAVLEGTAFGLRYGLDLRLASGLPGEIEIRLVGAA 414
QY 265 VFELMNRVVRVMYQTMQDQRLNLEPLPRILPIPGCGS--EAACSL 307
Db 415 KNPLMRR-----TLADLLGLFLVCPQTEAAALGAALQAANSL 452

RESULT 15
H70872
probable ctpD protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
R:Accession: H70872
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: H70872
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-657 <COL>
A:Cross-references: GB:AL021184; GB:AL123456; NID:93261498; PIDN:CAA15997.1; PID:927914
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: ctpD
C:Superfamily: Enterococcus copper-transporting ATPase copB; ATPase nucleotide-binding
F:25-394/Domain: ATPase transduction domain homology <ATT>
F:465-608/Domain: ATPase nucleotide-binding domain homology <ATN>

Query Match 6.9%; Score 113.5; DB 2; Length 657;
Best Local Similarity 20.4%; Pred. No. 0.099;
Matches 80; Conservative 50; Mismatches 125; Indels 137; Gaps 17;
QY 8 LLALAAAPAAAAEPLKLEKVVIL-----SRHGVR-----PTKALMVSP 48
Db 96 LLMIAAAGVAVAIQIFDGAALLIVFATSGALDDIATHTAESVKGLLDLAPDQAVVVG 155
QY 49 DWPWVGLTPRGAALVLLGYRYFARGLLPG--CPAAGTIL---ADVDERTRTGOFAAGL 103
Db 156 DGSERV-----VAASELVVGDVVVR---PGDRIPADGAVLSGASDVQDSITGESMPVAK 208
QY 104 APGCAIVHAGDDPIFHGLDTG-----CLDQADAILA-----AGEGGITARHGLTIA 149
Db 209 ARG-----DEVFAGTVNGSGVLHVYTRDPSTQTVVARIVELVADASATKATQLEFIE 260
QY 150 KV-----LNFASACLECARVVGRLGPLLASTL 176
Db 261 KIEQRYSLGMVAUTALIVILMFCADLRPVLLRAMTENVIVASPCAVVLATMPPLLSAIA 320
QY 177 SEIFLLEYAOGPMVEVGMGR-----ISAEMLLHNAQFLMNRTPYIARATPILIVTALS 228
Db 321 NA-----GRHGVLKSAVVVERLADTSIVALDKTGLTRGIP----- 357
QY 229 PARVVLLAGHDNTNALGGLDLWQL-----LPQDETPPGGALVFELMNRVVRVMYQTMQDQLR 284
Db 358 --RLASVAPLDPNV-VDARRLLQLAAAEQSEHPLGRAIVAEARRRGIAI----- 405
QY 285 NLEPLPRILPIPGCGSEAAACSLSDFARLVAP 316
Db 406 ---PPAKDFRAVPGCGVHALVG-NDFVEIASP 433

Search completed: April 30, 2004, 12:42:22
Job time : 13.6341 secs

OM protein - protein search, using sw model

Run on: April 30, 2004, 12:31:10 ; Search time 7.7561 Seconds
(without alignments)
2134.876 Million cell updates/sec

Title: US-10-021-723B-14
Perfect score: 1647
Sequence: 1 VSLLLAALLALALAPAAA.....CGSEAAACSLSDPARLVAPAC 318

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues
Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	613.5	37.2	432	1 PPA_ECOLI	P07102 escherichia
2	288.5	17.5	413	1 AGP_ECOLI	P19926 escherichia
3	284.5	17.3	413	1 AGP_SALTY	O33921 salmonella
4	256	15.5	417	1 AGP_PRORE	O52309 providencia
5	134.5	8.2	423	1 PPAL_RAT	P20611 rattus norv
6	127.5	7.7	423	1 PPAL_MOUSE	P24638 mus musculu
7	124.5	7.6	423	1 PPAL_HUMAN	P11117 homo sapien
8	113.5	6.9	657	1 CTDP_MYCTU	O53160 mycobacteri
9	103.5	6.3	386	1 PPAP_HUMAN	P15309 homo sapien
10	102.5	6.2	312	1 FMT_BORER	Q7WQ88 bordetella
11	102.5	6.2	312	1 FMT_BORFA	Q7WV20 bordetella
12	100.5	6.1	411	1 YOH2_STRCO	P40180 streptomyc
13	98.5	6.0	492	1 MOT3_RAT	O70461 rattus norv
14	98	6.0	402	1 KAS2_STRCM	P41176 streptomyc
15	94	5.7	1307	1 BCC4_ACEXY	O9WX71 acetobacter
16	92	5.6	1098	1 EMBB_MYCTU	P72030 mycobacteri
17	91.5	5.6	245	1 TRMD_CAUCR	Q9ABM5 caulobacter
18	91	5.5	772	1 METE_STRCO	Q93J59 streptomyc
19	90.5	5.5	352	1 TPO_CANFA	P42705 canis famil
20	90	5.5	669	1 PP1B_STRCO	O9KY04 streptomyc
21	89.5	5.4	312	1 FMT_BORPE	O7V889 bordetella
22	89	5.4	402	1 DEX_SYNLE	O9RGT1 synchococc
23	88.5	5.4	432	1 MESJ_ECOLI	P52037 escherichia
24	88.5	5.4	449	1 AROA_PSES2	P56952 pseudomonas
25	88.5	5.4	450	1 AROA_BRUME	Q9YEG1 bruceella me
26	88.5	5.4	450	1 AROA_BRUSE	Q8G3C4 bruceella su
27	88	5.3	776	1 HYPF_AZOVI	P40596 azotobacter
28	88	5.3	3011	1 POLG_HCVH	P27958 h genome po
29	87.5	5.3	640	1 Y051_MYCTU	O10880 mycobacteri
30	87.5	5.3	1164	1 YAV1_XANCV	P14727 xanthomonas
31	87.5	5.3	2554	1 TLES_DRONE	P13368 drosophila
32	87	5.3	375	1 COBT_STRAW	Q82AM7 streptomyc
33	87	5.3	405	1 YE01_PYRAB	Q9UYV6 pyrococcus

34 87 5.3 635 1 VP40_HSV11 P10210 herpes simp
35 87 5.3 1008 1 MMLA_MYCLE Q49619 mycobacteri
36 86.5 5.3 382 1 NORW_VIBVU Q844f7 vibrio vuln
37 86.5 5.3 503 1 PUR8_STRLP P42670 streptomyce
38 86.5 5.3 623 1 GLMS_MYCTU O06253 m glucosami
39 86 5.2 341 1 FAH1_DEIRA Q9T223 deinococcus
40 85.5 5.2 450 1 AROA_BRUAB Q9EGV2 bruceella ab
41 85 5.2 228 1 BIOD_PSEAE Q91614 pseudomonas
42 85 5.2 405 1 CHSY_EQUAR Q9MBB1 equisetum a
43 85 5.2 654 1 SPH2_HUMAN Q9NRA0 homo sapien
44 85 5.2 836 1 VG26_BPMD2 O64220 mycobacteri
45 84.5 5.1 418 1 CLCB_ECOLI P76175 escherichia

ALIGNMENTS

RESULT 1
PPA_ECOLI
ID_PPA_ECOLI STANDARD; PRT; 432 AA.
AC P07102;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Periplasmic appA protein precursor [Includes: Phosphoanhydride
DE phosphohydrolase (EC 3.1.3.2) (pH 2.5 acid phosphatase) (AP); 4-
DE phytase (EC 3.1.1.3.26)].
GN APPA OR B0980.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-29.
RC STRAIN=K12;
RX MEDLINE=90368616; PubMed=2168385;
RA Dassa J., Marck C., Boquet P.L.;
RT "The complete nucleotide sequence of the Escherichia coli gene appA
RT reveals significant homology between pH 2.5 acid phosphatase and
RT glucose-1-phosphatase."
RL J. Bacteriol. 172:5497-5500(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakano Y., Nishimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12,7-28.0 min region on the linkage map."
RL DNA Res. 3:1137-155(1996).
RN [4]
RP SEQUENCE OF 1-112 FROM N.A.
RX MEDLINE=87271766; PubMed=3038201;
RA Touati B., Danchin A.;
RT "The structure of the promoter and amino terminal region of the pH
RT 2.5 acid phosphatase structural gene (appA) of E. coli: a negative
RT control of transcription mediated by cyclic AMP."
RL Biochimie 69:215-221(1987).
RN [5]

Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
Sampai G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
Yano M., Horiuchi T.;
"A 718-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map.";
DNA Res. 3:137-155(1996).
[4]
SEQUENCE OF 23-34.
STRAIN=K12 / EMG2;
MEDLINE=97443975; PubMed=9298646;
Link A.J., Robison K., Church G.M.;
"Comparing the predicted and observed properties of proteins encoded
in the genome of Escherichia coli K-12.";
Electrophoresis 18:1259-1313(1997).
-!- FUNCTION: ABSOLUTELY REQUIRED FOR THE GROWTH OF E.COLI IN A HIGH-
PHOSPHATE MEDIUM CONTAINING G-1-P AS THE SOLE CARBON SOURCE.
-!- CATALYTIC ACTIVITY: D-glucose 1-phosphate + H(2)O = D-glucose +
phosphate.
-!- ENZYME REGULATION: INDEPENDENT FROM INORGANIC PHOSPHATE
AVAILABILITY, AND APPARENTLY SUBMITTED TO CATABOLITE REPRESSION.
IT IS POSITIVELY CONTROLLED BY CAMP AND THE CAMP RECEPTOR PROTEIN.
-!- SUBUNIT: Homodimer.
-!- SUBCELLULAR LOCATION: Periplasmic.
-!- MISCELLANEOUS: OPTIMALLY ACTIVE AROUND PH 4.
-!- SIMILARITY: Belongs to the histidine acid phosphatase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; M33807; AAA23426.1; -;
EMBL; AE000202; AAC74087.1; -;
EMBL; D90737; BAA35769.1; -;
EMBL; D90738; BAA35779.1; -;
PIR; JY0087; JY0087.
HSP; P07102; LDKM.
Ecogen; EGI0033; agp.
InterPro; IPR000560; HisAc_phosphatase.
Pfam; PF00328; acid_phosphat; 1.
PROSITE; PS00016; HIS_ACID_PHOSPHAT_1; 1.
PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
Hydrolase; Periplasmic; Signal; Complete proteome.
SIGNAL 1 22
FT CHAIN 23 413 GLUCOSE-1-PHOSPHATASE.
FT ACT SITE 40 40 NUCLEOPHYLIC ACCEPTOR (BY SIMILARITY).
FT ACT SITE 311 311 PROTON DONOR (BY SIMILARITY).
SO SEQUENCE 413 AA; 45683 MW; ADAD3363D0D6AB CRC64;

[illegible]

Db 179 DSYQLEKIVNYKDSACKKQOCSLVGKNTFSAKYQOEPGVSGPLKVGNSLVDAFTLQ 238
QY 184 YAQG-PM-EVWGGRISA-EW-----LLHNAQFLMNRTPYIAR--ATPI-----LIVT 225
Db 239 YVEGFPDQVANGKEIKSDQKWLKNGYQDSIFTSPEVARNVAKPLVSYIDKALVTD 298
QY 226 ALSPARVLLAGHDNLA--LGLDL--WQL-PQDETTPGGLVFLWM-----NR---Y 272
Db 299 RTSAPKITVLVGHDSNIALSLTALDFKPYQLHDQNERTPIGGKIVFQRMHDSKANRDLMK 358
QY 273 VRVYQTMQDLNLEPLPR-----ILPIQC--GSEAACSLSDF 310
Db 359 IYVYQSAEQLNADALTLQAPQAVTLELSGCPIDADGFCPMDKF 404

RESULT 3

AGP_SALTY
ID AGP_SALTY STANDARD; PRT; 413 AA.
AC O33921;
DT 15-JUL-1998 (Rel. 36, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glucose-1-phosphatase precursor (EC 3.1.3.10) (Glpase).
GN AGP OR STM117.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2.";
RL Nature 413:852-856(2001).
RN [2]
RP SEQUENCE OF 1-249 FROM N.A.
RC STRAIN=TN1379;
RX MEDLINE=97405892; PubMed=9260936;
RA Gupta S.D., Wu H.C., Rick P.D.;
RT "A Salmonella typhimurium genetic locus which confers copper
tolerance on copper-sensitive mutants of Escherichia coli.";
RL J. Bacteriol. 179:4977-4984(1997).
CC -|- CATALYTIC ACTIVITY: D-glucose 1-phosphate + H(2)O = D-glucose +
phosphate.
CC -|- SUBUNIT: Homodimer (By similarity).
CC -|- SUBCELLULAR LOCATION: Periplasmic.
CC -|- SIMILARITY: Belongs to the histidine acid phosphatase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; A3008748; AAL20049.1; -
DR EMBL; U75949; AAC45604.1; -
DR KSSP; P07102; IDKU.
DR ScyGene; SG10595; agp.
DR InterPro; IPR000560; HisAc_phsphtase.
DR Pfam; PF00328; acid_phosphat; 1.
DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
KW Hydrolase; Periplasmic; Signal; Complete proteome.
FT SIGNAL 1 22 BY SIMILARITY.
FT CHAIN 23 413 GLUCOSE-1-PHOSPHATASE.

FT ACT_SITE 40 40 NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).
FT ACT_SITE 311 311 PROTON DONOR (BY SIMILARITY).
FT CONFLICT 57 57 T -> P (IN REF. 2).
FT CONFLICT 60 63 AWPA -> TCPP (IN REF. 2).
FT CONFLICT 74 74 K -> T (IN REF. 2).
SQ SEQUENCE 413 AA; 45559 MW; E057667CF8A7244C CRC64;
Query Match 17.3%; Score 284.5; DB 1; Length 413;
Best Local Similarity 30.8%; Pred. No. 6.5e-15;
Matches 126; Conservative 59; Mismatches 127; Indels 97; Gaps 29;
QY 4 LLAALLAALAPAAAAA---EPLKLEKWLILSRHGVRSPT---KALMVS-----PDW 50
Db 5 LLAVAGAVLSSAVQACTTPEGYLOQLVMSRHLRLAPLANNGSVLAQSTPNAPAW 64
QY 51 FWPVG-LTPRGAL-VLLGYR-----YFARGLLPG--CPAAGTILADVDERTR---TGOAF 99
Db 65 DVEGGQLTTGGVLEVINGHYTREMLVAOGLIPSECPADPTVYVANSLOQTVAQFF 124
QY 100 AAGLAPGCAI-VHAGD-----DPIFHGLDTGCLDOADAILAAGEGGLTARHGI----- 146
Db 125 ITGAFPGCDIPVHHQKMGMTDPTFNPVTD--DSAPFQQAVQAMEKARSQHLDES YK 182
QY 147 TLAKVLNF--ASACLE--CARV-----VGRLGPI-LASTLSEIFLEIYAQG 187
Db 183 LLEQITHYQDSPCKEKHOCSLIDAKDTPSANYQOEPGVQGLKVGNSLVDAFTLQY YEG 242
QY 188 -PM-EVWGRI--SAEW-----LLHNAQFLMNRTPVIAR--ATP-----LIVTALSP 229
Db 243 FPDQVANGIHTDRQWKLKNGYQDSIFTSPTVARNVAPLKYIDKVLVADRVSA 302
QY 230 ARVLLAGHDNLA--LGLDL--WQL-PQDETTPGGLVFLWM-----NR---YVRVM 276
Db 303 PKYTVLVGHDSNIALSLTALDFKPYQLHDQYERTPIGGQLVQFQRMHDSKANRDLMKIE YV 362
QY 277 YQTMQDLNLEPLPR-----ILPIPGGSEA--ACSLSDARLAVAPA 317
Db 363 YQSARQLRNBALTLKSPAQRVTLKGGCPVDANGFCPLDKFDNVNNTA 411

RESULT 4

AGP_PROBE
ID AGP_PROBE STANDARD; PRT; 417 AA.
AC Q52309;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Glucose-1-phosphatase precursor (EC 3.1.3.10) (Glpase).
GN AGP.
OS Providencia rettgeri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Providencia.
OX NCBI_TaxID=587;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PV7;
RA Riccio M.L., Chiesurin A., Lombardi G., Satta G.;
RL Submitted (SEP-1992) to the EMBL/GenBank/DBJ databases.
CC -|- CATALYTIC ACTIVITY: D-glucose 1-phosphate + H(2)O = D-glucose +
phosphate.
CC -|- SUBUNIT: Homodimer (By similarity).
CC -|- SUBCELLULAR LOCATION: Periplasmic (By similarity).
CC -|- SIMILARITY: Belongs to the histidine acid phosphatase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X68201; CAA48288.1; -
DR

DR PIR; S25627; S25627.
DR HSP; P07102; 1DKL.
DR InterPro; IPR000560; HisAc_pheptase.
DR Pfam; PF00328; acid_phosphat; 1.
DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
KW Hydrolase; Periplasmic; Signal.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 417 GLUCOSE-1-PHOSPHATASE.
FT ACT SITE 42 42 NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).
FT ACT SITE 314 314 PROTON DONOR (BY SIMILARITY).
SQ SEQUENCE 417 AA; 46940 MW; 8CC4CA4F55DBF90 CRC64;

Query Match 15.5%; Score 256; DB 1; Length 417;
Best Local Similarity 29.0%; Pred. No. 1.1e-12; Indels 94; Gaps 27;
Matches 119; Conservative 57; Mismatches 140;

QY 1 VSLLAALLAALAPAAAAE---PLKLEKVLISRHGRVSP--TKALMVS-----PD 49
DB 6 LTLCLSAAL-FAPAPTMASTDNQADMYLDQVLVSRHNLRTPIVNTGILTEVTDKQWP 64

QY 50 WPNVPG-LTPGAALVILGYFARGLLPG-----CPAAGT---TLADVDEPT-RTGQ 97
DB 65 WDAKSGYLTTGGALEVYMGFWIDQNKLLADELCPTSNEDIYLTNSLQRTIATAQ 124

QY 98 AFAAGLAPGCAI-VH-----AGDDPIFHGLDT-GCLQADAILAAGEG---GLTARHGL- 146
DB 125 FFAAGAPGCKVNIHQBEICKMDPVFNPIITNSPBEKQKALAAADDYLGSLKAGYE 184

QY 147 TLAKVLNFA-----SACLECARVVRGLGPI-LASTLSIEIFLEVAQG 187
DB 185 ELDTVNLKDSQKCKTKLCLNLDSCQKNSFIIEADKEFGVSGPLKIANSAVDAIDLQYEG 244

QY 188 --PMEVHGGRISA-EW-----LLHNAQFLMNRTPVIAR---ATPIL-----ITVALS 228
DB 245 FPAQVANGLVDTKEKWKLNLTAKNAVQETLFTKILAKNVAHFILNYIDKGFVVDKGE 304

QY 229 PARVVLGAGHTNLA--LGGDL--WQLPQ-DETPPGGALVFLMN-----RVRV 275
DB 305 TAKTIFLVGHSNIAASLMSADFPYQLAQOQYETPIGKLVFORWTDKTKDFMKVEY 364

QY 276 MYQTMQDLR-----NLE-PLPLPRILPIPGC--GSEACSLSDFARLVAPA 317
DB 365 VIQTADQRDNAYLSLETTPPKHVTLELDCPVDKNGKCSWEDFOKWAKA 414

RESULT 5
ID PPAL RAT STANDARD; PRT; 423 AA.
AC P20611.
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Lysosomal acid phosphatase precursor (EC 3.1.3.2) (LAP).
GN AC2P.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89350910; PubMed=2764916;
RA Himeno M., Fujita H., Noguchi Y., Kono A., Kato K.;
RT "Isolation and sequencing of a cDNA clone encoding acid phosphatase
in rat liver lysosomes."
RL Biochem. Biophys. Res. Commun. 162:1044-1053(1989).
CC -!- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O = an
alcohol + phosphate.
CC -!- SUBCELLULAR LOCATION: Lysosomal.
CC -!- SIMILARITY: Belongs to the histidine acid phosphatase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
EMBL; M27893; AAA40744.1; -
PIR; A33395; A33395.
DR HSP; P20646; IRPA.
DR InterPro; IPR000560; HisAc_pheptase.
DR Pfam; PF00328; acid_phosphat; 1.
DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
KW Hydrolase; Signal; Glycoprotein; Lysosome.
FT SIGNAL 1 30
FT CHAIN 31 423 LYSOSOMAL ACID PHOSPHATASE.
FT DISULFID 159 370 BY SIMILARITY.
FT DISULFID 212 310 BY SIMILARITY.
FT DISULFID 345 349 BY SIMILARITY.
FT ACT SITE 42 42 NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).
FT ACT SITE 84 84 BY SIMILARITY.
FT ACT SITE 286 286 PROTON DONOR (BY SIMILARITY).
FT CARBOHYD 92 92 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 167 167 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 177 177 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 197 197 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 267 267 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 423 AA; 48319 MW; 6F4C4819CADD4496 CRC64;

Query Match 8.2%; Score 134.5; DB 1; Length 423;
Best Local Similarity 25.3%; Pred. No. 0.0029;
Matches 95; Conservative 33; Mismatches 157; Indels 91; Gaps 20;

QY 4 LLAALLAALLAALAPAAAAEPLKLEKVLISRHGRVSPKALMVSP--DWPWVG---LTP 58
DB 13 LLQFLGMLVMVPPIQARSLRF--VTLRYHGDSRPVKAYPKDPYQBEKWPQGFQLTK 70

QY 59 RGAAL-----VLGVRYFARGLLPGCPAAGTIL---ADVDERTRTGOAPAGLAPCAIV 110
DB 71 EGMQLHWELGQALRQRY--HGFNLASYHQVYVSTDFDRLMSABANLAGLFPFTEVQ 128

QY 111 H-----AGDDPIFHGLDTG-----CLDQADAI-L 133
DB 129 HFNPNISWQPIPVHTVPITDELLKFPGLGCPRYEQQLQNETRQTPYQNMISQNAQFLDM 188

QY 134 AAGEGLTARHGLTLAKVLNFAACLECARVVRGLGPIAS-----TLSEI---FLLEY 184
DB 189 VANETGLM--NLLETINWVDT-LFCEQTHGLLPPWASPTQVQALSQDKDFSLFLP 244

QY 185 AQCPMEVGMGRISAELWLLHNAQFLMNRTPYIARATPILIVTALS PARVVLLAGHDTNLA 244
DB 245 GIHD-QVQKARLQGGVLL--AQILKNLT-----LMATTSPFKLLIVYSAHDTTLVA 292

QY 245 GGLDLWQLPQDETPPGGALVFLMN-----NRYRVVMYQTMQDLRNLPLPRILP 300
DB 293 LQNAL--NVYNGKQAPYASCHIFEOYDNGNGNSVEMY-----FRNDSKAPMLPTLPGCP 346
301 SEAAACSLSDFARLVAP 316
347 HR--CPQLQDFLRLTEP 360

RESULT 6
ID PPAL MOUSE STANDARD; PRT; 423 AA.
AC P24638; Q8QZT5;
DT 01-MAR-1992 (Rel. 21, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE GN ACP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Liver;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K.B., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Lquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richardson S., Worley K.C., Hale S.E., Garcia A.M., Gay L.J., Rulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RN SEQUENCE OF 3-423 FROM N.A.
RX MEDLINE=91282986; PubMed=2059337;
RA Geier C., von Figura K., Pohlmann R.;
RT "Molecular cloning of the mouse lysosomal acid phosphatase.";
RL Bio. Chem. Hoppe-Seyler 372:301-304(1991).
CC -|- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O = an
alcohol + phosphate.
CC -|- SUBCELLULAR LOCATION: Lysosomal.
CC -|- SIMILARITY: Belongs to the histidine acid phosphatase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
DR EMBL; BC023343; AAH23343.1; -;
DR EMBL; X57199; CAA040485.1; -;
DR F1R; S14742; S14742.
DR HSP; P15309; 2HPA.
DR MGD; MGI:87882; ACP2.
DR InterPro: IPR000560; HisAc_phsphtase.
DR Pfam: PF00328; acid_phosphat; 1.
DR PROSITE; PS00616; HIS_ACID_PHOSPHAT 1; 1.
DR PROSITE; PS00778; HIS_ACID_PHOSPHAT 2; 1.
KW Hydroxylase; Signal; Glycoprotein; Lysosome.
FT SIGNAL 1 30
FT CHAIN 31 423 LYSOSOMAL ACID PHOSPHATASE.
FT DISULFID 159 370 BY SIMILARITY.
FT DISULFID 212 310 BY SIMILARITY.
FT DISULFID 345 349 BY SIMILARITY.
FT ACT SITE 42 42 NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).
FT ACT SITE 84 84 BY SIMILARITY.
FT ACT SITE 286 286 PROTON DONOR (BY SIMILARITY).
FT CARBOHYD 92 92 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 133 133 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 167 167 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 177 177 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 191 191 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 197 197 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 267 267 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 322 322 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 331 331 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 374 374 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 131 131 N -> S (IN REF. 2).
SQ SEQUENCE 423 AA; 48508 MW; AFD7ASC90A4F2AF4 CRC64;
Query Match 7.7%; Score 127.5; DB 1; Length 423;
Best Local Similarity 23.5%; Pred. No. 0.01;
Matches 92; Conservative 32; Mismatches 145; Indels 123; Gaps 19;
QY 6 AALL-----ALAAAPAAAAAEPLKLVILSRHVSPTKALMVSP-----DWP----- 51
DB 11 AALLQFLGCLTVMPIQARS-----LRFTLLYRHGDRSPVKTYPKDYQBEKMPQGG 66
QY 52 -----WPVGLTPRGAALVLLGYR-----FARGLLPGCPAAGTTLADVDERTRG 96
DB 67 QLTKEGMLQHWELGQALRQRYHGFLNTSVHRQEVYRS-----TDFDRLMSA 114
QY 97 QAFAGIAPGCAIVH-----AGDDIFHGL----- 121
DB 115 EANLAGLFPNEVQHFNPNISWQIPVHTVPTEDRLKPLGCPRYFOLQNETQTPE 174
QY 122 -DTGCLDQDAI-LAAGEGLTARHGLTLAKVILNFASACLECARVVGRGLPLIAS----- 174
DB 175 YQNRSTQNAQFLNMANETGLT-----NVLTETINWVYDT-LFCEQTHGLLLPPWASPQTQV 230
QY 175 -----TLSEIFLLEVAQSPMEVWGRIASAEWLLHNAQFLMRTPIARATPILIVTALS 228
DB 231 RLSQLKDFSLFLFGIHE---QVKARLQGGVLL---AQILKNLT-----LMATTSQ 276
QY 229 PARVILLAGHTNLALGGLDLWOLPQDPTTPPGALVFLW-----NRYVVMYQTMDQLR 284
DB 277 FPKLLVYSAHDTTLVALQMAL-NVYNGKQAPYASCHIFELYQEDNGNFSVMY-----FR 330
QY 285 NLSEFLPLRILPIPGCGSEAACSLSDFARLVAP 316
DB 331 NDSKQAPWPLILPGCPHR--CPLQDFLRLETP 360
RESULT 7
PPAL_HUMAN STANDARD; PRT; 423 AA.
ID AC P1117; Q9BTU7;
DT 01-JUL-1989 (Rel. 11, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Lysosomal acid phosphatase precursor (EC 3.1.3.2) (LAP).
GN ACP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Placenta;
RX MEDLINE=89052645; PubMed=3191910;
RA Pohlmann R., Krentler C., Schmidt B., Schroeder W., Lorkowski G.,
RA Cullley J., Merseman G., Geier C., Waheed A., Gottschalk S.,
RA Grzeschik K.H., Hasikik A., von Figura K.;
RT "Human lysosomal acid phosphatase: cloning, expression and
chromosomal assignment.";
RL EMBL J. 7:2343-2350(1988).
RN [2]
RN SEQUENCE FROM N.A.
RP TISSUE=Leukocyte;
RX MEDLINE=89377828; PubMed=2776754;
RA Geier C., von Figura K., Pohlmann R.;
RT "Structure of the human lysosomal acid phosphatase gene.";
RL Eur. J. Biochem. 183:611-616(1989).
RN [3]
RN SEQUENCE FROM N.A.
RC TISSUE=Kidney;

Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A., Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.; "Whole-genome comparison of Mycobacterium tuberculosis clinical and laboratory strains."; J. Bacteriol. 184:5479-5490(2002).

[3] SEQUENCE FROM N.A.

RRP SPECIES=M.bovis; STRAIN=AF2122/97;

RC MEDLINE=22709107; PubMed=12788972;

RRX Garnier T., Eglmeier K., Camus J.-C., Medina N., Mansoor H., Pryor M., Duthoy S., Grondin S., Lacroix C., Monsenpe C., Simon S., Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R., Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.; "The complete genome sequence of Mycobacterium bovis."; Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).

CCC -!- CATALYTIC ACTIVITY: ATP + H(2O) = ADP + phosphate.

CCC -!- SUBCELLULAR LOCATION: Integral membrane protein.

CCC -!- SIMILARITY: LOCATIONS to the cation transport ATPases family (P-type ATPases). Subfamily IB.

CCC -----

CCC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CCC -----

CCC EMBL; AL021184; CAA15997.1; -

DR EMBL; AE007021; AAK45780.1; -

DR EMBL; BX248339; CAD96171.1; -

DR PIR; H70872; H70872.

DR TIGR; MT1515; -

DR Tuberculist; Rv1469; -

DR InterPro; IPR006416; ATPase-IB_hvy.

DR InterPro; IPR001757; ATPase-IB_hvy.

DR InterPro; IPR001366; Cad ATPase.

DR InterPro; IPR008250; E1-E2 ATPase reg.

DR InterPro; IPR006404; Heavy met ATPase.

DR InterPro; IPR005834; Hydrolase.

DR Pfam; PF00122; E1-E2 ATPase; 1.

DR Pfam; PF00702; Hydrolase; 1.

DR PRINTS; PR00119; CATATPASE.

DR PRINTS; PR00941; COATPASE.

DR TIGRFAMS; TIGR01512; ATPase-IB2_Cd; 1.

DR TIGRFAMS; TIGR01525; ATPase-IB_hvy; 1.

DR TIGRFAMS; TIGR01494; ATPase_P-type; 2.

DR PROSITE; PS00154; ATPASE E1_E2; 1.

DR Hydrolase; Transmembrane; Phosphorylation; Magnesium; ATP-binding; Complete proteome.

KW Complete proteome.

KW TRANSMEM 40 60 POTENTIAL.

FT TRANSMEM 62 82 POTENTIAL.

FT TRANSMEM 102 122 POTENTIAL.

FT TRANSMEM 268 288 POTENTIAL.

FT TRANSMEM 299 319 POTENTIAL.

FT TRANSMEM 447 467 POTENTIAL.

FT TRANSMEM 477 497 POTENTIAL.

FT TRANSMEM 607 627 POTENTIAL.

FT TRANSMEM 347 347 PHOSPHORYLATION (BY SIMILARITY).

FT MOD_RES 347 347 MAGNESIUM (BY SIMILARITY).

FT METAL 543 543 MAGNESIUM (BY SIMILARITY).

FT METAL 547 547 MAGNESIUM (BY SIMILARITY).

CC SEQUENCE 657 AA; 67884 MW; 1AF5A7DD4BC697D5 CRC64;

CC -----

Query Match 6.9%; Score 113.5; DB 1; Length 657;

Best Local Similarity 20.4%; Pres. No. 0.19;

Matches 80; Conservative

QY 8 LLALALAPAAAAAEPLKLEKVVIL-----SRHGVR-----PTKALMVSP 48

Db 96 LLMIAAVGAVGICQIPDGALLIVFATSGALDDIATRHFAESVKGLLDIAPQAVVQ 155

QY 49 DWFPVGLTPRGALVLLGYRFARGLLPG---CPAAGTIL---ADVDETRTQGAAGL 103


```
QY 99 FAAGLAP--CCAI-----VHA---GDP-----PIFGHLDTCGLDQ----- 128
Db 119 NLAAFPPEGVSWNPILLWQPIPVHTVPLSEQLLYLPFRNCPFELESETLKSEEFQ 178
QY 129 -----ADATLAAGEGLTARHGTLAKVLNFPASACLECARVVGRLGPLLAS----- 174
Db 179 KRLHPYKDFATLGGK--LSLGHQDLFGIWSKYVDPLYCESVHNFTLPSWATEDMTKLR 236
QY 175 TLSEIFLELYAQPMGVGWRISABWLLHNAQFLMRTPTPIARATPILIVTALSAPRVVL 234
Db 237 ELSELSLSLYGHGKHQKESRLOGGLVNE---ILN---HMKRATQI-----PSYKJLIM 285
QY 235 LAGHTNIALGGDLWLQLPQDPETPPGCA-----LVFELMNRVVRVYQTMQDLRLNLEPL 289
Db 286 YSAHDITVS--GLQVLDVYNGLLPPVASCHELTIFYEKEGYFVEMY-----RNETQH 337
QY 290 LPRILPIPGCGSZAACSLSPFARLVAP 316
Db 338 EYPLMLPGC--SPSCPLERFAELVGP 362

RESULT 10
FMT_BORBR
ID FMT_BORBR STANDARD; PRT; 312 AA.
AC Q7WQSS;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Methionyl-tRNA formyltransferase (EC 2.1.2.9).
GN FMT OR B02048.
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RB50 / ATCC BAA-588;
RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cardeno-Tarraga A.M., Temple L., James K., Bason N., Cherevach I.,
RA Achtman M., Atkin R., Baker S., Basham D., Harsnis B., Quail M.A.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Leather S., Moutle S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders R., Squares S., Stevens K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RA "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
CC -!- FUNCTION: Modify the free amino group of the aminoacyl moiety of
CC methionyl-tRNA(fMet). The formyl group appears to play a dual role
CC in the initiator identity of N-formylmethionyl-tRNA by: (i)
CC promoting its recognition by IP2 and (ii) impairing its binding to
CC EFTu-GTP (By similarity).
CC -!- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + L-methionyl-
CC tRNA(fMet) + H(2)O = tetrahydrofolate + N-formylmethionyl-
CC tRNA(fMet).
CC -!- SIMILARITY: Belongs to the fnt family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC ENBL; BX640437; CAB30746.1; -.
CC HAMAP; MF 00182; -.
DR InterPro; IPR005794; Fmt.
```

```
DR InterPro; IPR002376; formyl_transf.
DR InterPro; IPR005793; formyl_transf_C.
DR Pfam; PF00551; formyl_transf; 1.
DR Pfam; PF02911; formyl_transf_C; 1.
DR TIGRFAMs; TIGR00460; fnt; 1.
DR PROSITE; PS00373; GART; FALSE NEG.
KW Transferase; Methyltransferase; Protein biosynthesis;
KW Complete proteome.
FT BINDING 117..120 TETRAHYDROFOLATE (THF) (BY SIMILARITY).
SQ SEQUENCE 312 AA; 32601 MW; 1F70614C084D7B55 CRC64;

Query Match 6.2%; Score 102.5; DB 1; Length 312;
Best Local Similarity 25.4%; Pred. No. 0.64;
Matches 52; Conservative 24; Mismatches 66; Indels 63; Gaps 10;

QY 9 LALAALAPAAARAEPLKLEK-----VWLSRHGVSPFKALMVSPDWMPV----- 54
Db 65 LRLDGRYPDEAAARAQALRVAPDMVMVYAG-----LILQWTLDFRLGLNIHA 116
QY 55 GLTP--RGAALVILGVYFARGLLPGCPAAGTILADYDERTTG-----QAFAGLAPGC 107
Db 117 SLLPWRGAAP-----QRAIEAGDAETGVITMQMDAGLDTGDMLERAVPICAQQA 169
QY 108 AIVH-----AGDDPIFHGLDTGCLDQADAILAAGEGGITARH-----GLTLAKVLNFSAC 158
Db 170 AQLHDELALAGGQAI-----VDALAALGQGLAPRRQPDAGVTVAAKLDKAA 218
QY 159 LEC-----ARVVGRLGPLLAATL 176
Db 219 LDCSLPRAAVLARRVRAFNVPVGAII 243

RESULT 11
FMT_BORBR
ID FMT_BORBR STANDARD; PRT; 312 AA.
AC Q7WIV2;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Methionyl-tRNA formyltransferase (EC 2.1.2.9).
GN FMT OR B02048.
OS Bordetella parapertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=519;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=12822 / ATCC BAA-587;
RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cardeno-Tarraga A.M., Temple L., James K., Bason N., Cherevach I.,
RA Achtman M., Atkin R., Baker S., Basham D., Harsnis B., Quail M.A.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holtroyd S., Jagels K.,
RA Leather S., Moutle S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders R., Squares S., Stevens K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RA "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
CC -!- FUNCTION: Modify the free amino group of the aminoacyl moiety of
CC methionyl-tRNA(fMet). The formyl group appears to play a dual role
CC in the initiator identity of N-formylmethionyl-tRNA by: (i)
CC promoting its recognition by IP2 and (ii) impairing its binding to
CC EFTu-GTP (By similarity).
CC -!- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + L-methionyl-
CC tRNA(fMet) + H(2)O = tetrahydrofolate + N-formylmethionyl-
CC tRNA(fMet).
CC -!- SIMILARITY: Belongs to the fnt family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```


CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; BX640423; CAE39985.1; AUT_INIT.

DR HAMAP; MF 00182; -; 1.

DR InterPro; IPR005794; Fmt.

DR InterPro; IPR002376; formyl_transf.

DR Pfam; PF00551; formyl_transf; 1.

DR Pfam; PF02911; formyl_transf; 1.

DR TIGRFams; TIGR00460; fnt; 1.

DR PROSITE; PS00373; GART; FALSE NEG.

KW Transferase; Methyltransferase; Protein biosynthesis;

KW Complete proteome.

FT BINDING 117 TETRAHYDROFOLATE (THF) (BY SIMILARITY).

SQ SEQUENCE 312 AA; 32529 MW; 1F7060A6084D7BDF CRC64;

Query Match 6.2%; Score 102.5; DB 1; Length 312;

Best Local Similarity 25.4%; Pred. No. 0.64;

Matches 52; Conservative 24; Mismatches 56; Indels 53; Gaps 10;

QY 9 LALAALAPAAAAAEPLKLEK-----WILSRHGVSPTKALMVSPDWPV-----54

DB 65 LRLDGRYPDEAAAAAARACLERVAPDVMVVAAYG-----LILPQWTLPLRLGCLNIHA 116

QY 55 GLTP--RGAALLVGLYFARGLLPGCPAGTILADVDRTRTG-----QAFAGLAPGC 107

DB 117 SLPRWRGARI-----QRAIAGDAETGVITQMDAGLDGMDLLERAVFIGAQOTA 169

QY 108 AIVH-----AGDDPIFHGLDTCGLDQADAILAAGEGLTARH-----GLTLAKVLNIFASAC 158

DB 170 AQLHDELALAGQAI-----VDALAALGGGLAPRRQPDAGVTVAAKLDKAEAA 218

QY 159 LEC-----ARVVGRLGPLLSTL 176

DB 219 LDCSLPAAVLARRVRAFNFPVGAT 243

RESULT 12

ID YOH2 STRCO STANDARD; PRT; 411 AA.

AC P40180; Q9S6U5;

DT 01-FEB-1995 (Rel. 31, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Putative glycosyl transferase SC03672 (BC 2.-.-.-).

GN SC03672 OR SCH35.52.

OS Streptomyces coelicolor

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Streptomycinae; Streptomycetaceae; Streptomyces.

OX NCBI_TaxID:1902;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);

EX MEDLINE=96293864; PubMed=8722574;

RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 RL Nature 417:141-147 (2002).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: Belongs to the glycosyltransferase family 4.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; X77458; CAAS4605.1; -.

DR EMBL; AL939117; CAB44375.1; -.

DR FIR; S41945; S41945.

DR InterPro; IPR00715; Glyco trans 4.

DR Pfam; PF00953; Glycos transf 4; 1.

KW Hypothetical protein; Transferase; Transmembrane; Complete proteome.

FT TRANSMEM 7 27 POTENTIAL.

FT TRANSMEM 45 65 POTENTIAL.

FT TRANSMEM 70 90 POTENTIAL.

FT TRANSMEM 120 140 POTENTIAL.

FT TRANSMEM 148 168 POTENTIAL.

FT TRANSMEM 169 189 POTENTIAL.

FT TRANSMEM 197 217 POTENTIAL.

FT TRANSMEM 227 247 POTENTIAL.

FT TRANSMEM 277 297 POTENTIAL.

FT TRANSMEM 301 321 POTENTIAL.

FT CONFLICT 195 195 A -> R (IN REF. 1).

SQ SEQUENCE 411 AA; 41127 MW; A013111F7BC1CA9B CRC64;

Query Match 6.1%; Score 100.5; DB 1; Length 411;

Best Local Similarity 25.5%; Pred. No. 1.2;

Matches 74; Conservative 31; Mismatches 98; Indels 87; Gaps 15;

QY 2 SLLAALLAALAPAAAAAEPLKLEKVVLSRHGVSPTKALMVSPDWPVGLTPRGA 61

DB 17 SFLLTAVLAALLRAPAL-----RLAVDRRRRR-----EVPLL-GGV 52

QY 62 ALVLL-----GYRFARGLLPGCPAGTII-----ADVDETRTGQAFAG 102

DB 53 AVVLVTAVVAVAGDR---AGVVLPGAPAGRLLAATVVGALGLAADVWELRRLLAGTA 109

QY 103 LAPGCAIVHAGDDPIFHGLDTC-----LPQADAILAAGEGLTARHGLTLAKYL 152

DB 110 VRAACVVPYGETGPVAGALAVGWALVTGAPRGLDHADGV--GTVGVVTAFGVG-----162

QY 153 NFASACLECARVVGRLGP-----LLASTLSEIFLLEYAQGPMEVG-WGRISAENLLHNAQ 206

DB 163 -----ACAAVELMDGPVALLVLAALAGFLHNHHPARIALGACGLFTGFLLTGAA 215

QY 207 FLNMRTPY--IARATPILITVLTALSPARVVL-----LAGHDTNLALGLD 248

DB 216 VLA-RTGYGVGGAGVLCALTAVPVADAVLLLSRLAGRP--LSRGGPD 262

RESULT 13

MOT3 RAT

ID MOT3 RAT STANDARD; PRT; 492 AA.

AC 070461;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Monocarboxylate transporter 3 (MCT 3).

GN SLC16A8 OR MCT3.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Retina;
RX MEDLINE=99020931; PubMed=9841555;
RA Philip N.J., Yoon H., Grollman E.F.;
RT "Monocarboxylate transporter MCT1 is located in the apical membrane
of MCT3 in the basal membrane of rat RPE.";
RL Am. J. Physiol. 274:R1824-R1828 (1998).
CC -!- FUNCTION: Proton-linked monocarboxylate transporter. Catalyzes the
rapid transport across the plasma membrane of many
monocarboxylates such as lactate, pyruvate, branched-chain oxo
acids derived from leucine, valine and isoleucine, and the ketone
bodies acetoacetate, beta-hydroxybutyrate and acetate (By
similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Plasma membrane.
CC -!- TISSUE SPECIFICITY: RETINAL PIGMENT EPITHELIUM. RESTRICTED TO THE
BASOLATERAL MEMBRANE.
CC -!- SIMILARITY: Belongs to the major facilitator superfamily.
CC Monocarboxylate porter (TC 2.A.1.13) family.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; AF059258; AAC18120.1; --
DR InterPro; IPR007114; MFS.
DR InterPro; IPR004743; Monocarb transpt.
DR TIGRFAMs; TIGR00892; 2A0113; 1.
DR PROSITE; P50850; MFS; 1.
KW Transport; Symport; Transmembrane; Multigene family.
FT DOMAIN 1 14 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 15 35 POTENTIAL.
FT DOMAIN 36 58 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 59 79 POTENTIAL.
FT DOMAIN 80 85 CYTOPLASMIC (POTENTIAL).
FT TRANSNEM 86 106 POTENTIAL.
FT DOMAIN 107 115 EXTRACELLULAR (POTENTIAL).
FT TRANSNEM 116 136 POTENTIAL.
FT DOMAIN 137 147 CYTOPLASMIC (POTENTIAL).
FT TRANSNEM 148 168 POTENTIAL.
FT DOMAIN 169 172 EXTRACELLULAR (POTENTIAL).
FT TRANSNEM 173 193 POTENTIAL.
FT DOMAIN 194 230 CYTOPLASMIC (POTENTIAL).
FT TRANSNEM 231 251 POTENTIAL.
FT DOMAIN 252 257 EXTRACELLULAR (POTENTIAL).
FT TRANSNEM 258 278 POTENTIAL.
FT DOMAIN 279 293 CYTOPLASMIC (POTENTIAL).
FT TRANSNEM 294 314 POTENTIAL.
FT DOMAIN 315 318 EXTRACELLULAR (POTENTIAL).
FT TRANSNEM 319 339 POTENTIAL.
FT DOMAIN 340 352 CYTOPLASMIC (POTENTIAL).
FT TRANSNEM 353 373 POTENTIAL.
FT DOMAIN 374 386 EXTRACELLULAR (POTENTIAL).
FT TRANSNEM 387 407 POTENTIAL.
FT DOMAIN 408 492 CYTOPLASMIC (POTENTIAL).
FT TRANSNEM 492 AA; 51590 MW; 5CCC4DC3835FA0 CRC64;
SQ SEQUENCE 492 AA; 51590 MW; 5CCC4DC3835FA0 CRC64;

Query Match 6.0%; Score 98.5; DB 1; Length 492;
Best Local Similarity 25.5%; Pred. No. 2.1;
Matches 79; Conservative 29; Mismatches 123; Indels 79; Gaps 16;
QY 48 PD--WPPVGLTPRGAALVILGYRYPARGLLPGCP--AAGTILADVDTTGTGAFAAGLA 104
DB 13 PDGGWGVV-----LGACFVITGAY-----GFPKAVSVFFRELKEDFGAGYSDTAWS 61
QY 105 PGCAIVHAGDDPIFHGLDT--GCLDQADAILAGEGGLTARHGLTAKVLNFAACLE--160
DB 62 SIMLAMYGTGPISSILVTRFGC---RPVMLA---GGLLASAGMILA---SFASRLLELY 112

QY 161 -CARVVVGRGLGLASTLSEIFLLEYAGPMEVGMGRISAEWLLHNAQLMNRTPYIARAT 219
DB 113 LTAGVLTGLGLALNFQPSLIMGLYFERRPLANGLA--GS 153
QY 220 PIIIVTALSAPRVL---LAGHDTNIALGGLDL-----WQIQQPDETTPGG- 262
DB 154 PVELST--LSPGLQLGERFGWRGFLFLGGLLHCCACGAVMPPFPQPRPDAPPGR 212
QY 263 -----ALVPFLWRYVVMYQTDQLRNLEPLLPRL-PIPGGSEAAACSLSD 309
DB 213 ARHQQLDLAVCTRTFMVNTKFLMALGLFVAILLVNAYKADGVPDAAAFLLSIVG 272
QY 310 FARLVA-PAC 318
DB 273 FVDIVAREPAC 282
RESULT 14
KAS2 STRCM STANDARD; PRT; 402 AA.
ID KAS2 STRCM
AC P41176;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative polyketide beta-ketoacyl synthase 2 (EC 2.3.1.-) (ORF 2).
OS Streptomyces cinnamonensis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1900;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3823.5;
RX MEDLINE=92374994; PubMed=1508151;
RA Arrowmith T.J., Malpartida F., Sherman D.H., Hopwood D.A.,
RT Robinson J.A.;
RT "Characterisation of actI-homologous DNA encoding polyketide synthase
genes from the monensin producer Streptomyces cinnamonensis.";
Mol. Gen. Genet. 234:254-264 (1992).
CC -!- PATHWAY: Polyketide antibiotic monensin biosynthesis.
CC -!- MISCELLANEOUS: This putative ketoacyl synthase lacks the active
site cysteine.
CC -!- SIMILARITY: Belongs to the beta-ketoacyl-ACP synthases family.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; Z11511; CAA77597.1; --
DR PIR; S25077; S25077.
DR HSRP; P73283; 185M.
DR InterPro; IPR000794; Ketoacyl_synth.
DR Pfam; PF00109; Ketoacyl-synt_1.
DR Pfam; PF02801; Ketoacyl-synt_C; 1.
KW Antibiotic biosynthesis; Transferase; Acyltransferase.
SQ SEQUENCE 402 AA; 41270 MW; 348657D10201C214 CRC64;

Query Match 6.0%; Score 98; DB 1; Length 402;
Best Local Similarity 21.8%; Pred. No. 1.8; Indels 78; Gaps 12;
Matches 69; Conservative 37; Mismatches 133;
QY 23 FLKLEKVVILSRHGRVSPFKALMVSPPWMPVGLTPRGAALVLL-----GYRYPARGLLP 77
DB 3 PVAVTGMGIAAPNGILGRPTTG---RPPW-----APRAASAATRRFPDPSGYPQALAGEIP 53
QY 78 GCPRA-----GTLADVDTTGTGAFAAGLAPGCAIVHAGDDPIFHGLDTGCLDQADAIL 133
DB 54 GFRAEHLPGRLVPTDRVTRLSLAADWALADAGVEVAEDP-----LDWG-----VVT 103

QY 134 AAGCGGLTARHGLTLAKVLNFASACLECARVGRGLPGLLASTLSEIFLLEYAQGPMEVGM 193
 Db 104 ASHAGG-----FEQDELOKLLGQGPVLSAYOSFANFVAVNSGQISIRH 149
 QY 194 GRISAEWLHNAQ-----FLMNRTPYIARATPILIVTALSAPARVLLAGHDTNLAGGLD 248
 Db 150 GKGSPGVVYDQAGGLDALAQAARLRVKGTPLVICGAVEP-RSAPGAGSPSPA-GGMS 207
 QY 249 LWOLPQDETP-----PGGALVFELWNRYRVVYQTMQDLRLNLEPLLPRILPIGC 299
 Db 208 DSEPNRAYLPFRDGRGYVPGG-----RGVVPFLERAEARPAR 247
 QY 300 GSEACSLSDPARLVAP 316
 Db 248 GAEVYGEAGLRLAP 264

RESULT 15

BC4 ACXY STANDARD; PRT; 1307 AA.
 AC Q9WX71;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Cellulose synthase 2 operon protein C precursor.
 GN BCSCII.
 OS Acetobacter xylinus.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
 OC Acetobacteraceae; Gluconacetobacter.
 OX NCBI_TaxID=28448;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=JCM 7664 / IFO 13693;
 RX MEDLINE=99310341; PubMed=10382968;
 RA Umeda Y., Hirano A., Ishibashi M., Akiyama H., Onizuka T., Ikeuchi M.,
 RA Inoue Y.;
 RT "Cloning of cellulose synthase genes from Acetobacter xylinum JCM
 7664: implication of a novel set of cellulose synthase genes.";
 RL DNA Res. 6:109-115(1999).
 CC -!- FUNCTION: Required for maximal bacterial cellulose synthesis (By
 similarity).
 CC -!- PATHWAY: Bacterial cellulose biosynthesis.
 CC -!- SUBCELLULAR LOCATION: Outer membrane (Potential).
 CC -!- SIMILARITY: Belongs to the acsC/bcsc family.
 CC -!- SIMILARITY: Contains 9 TPR repeats.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC EMBL; AB015803; BAA77596.1; -
 DR InterPro; IPR008410; BCSC_C.
 DR InterPro; IPR003921; Cell_synth_C.
 DR InterPro; IPR008941; TPR-like.
 DR InterPro; IPR001440; TPR.
 DR Pfam; PF05420; BCSC_C; 1.
 DR Pfam; PF00515; TPR; 5.
 DR SMART; PR01441; CELLNTHASEC.
 DR SMART; SM00028; TPR; 4.
 KW Cellulose biosynthesis; Repeat; TPR repeat; Membrane; Outer membrane;
 Signal.
 FT SIGNAL. 1 55 POTENTIAL.
 FT CHAIN 56 1307 CELLULOSE SYNTHASE 2 OPERON PROTEIN C.
 FT REPEAT 97 130 TPR 1.
 FT REPEAT 270 303 TPR 2.
 FT REPEAT 339 372 TPR 3.
 FT REPEAT 374 406 TPR 4.
 FT REPEAT 458 491 TPR 5.
 FT REPEAT 493 525 TPR 6.

FT REPEAT 528 561 TPR 7.
 FT REPEAT 754 787 TPR 8.
 FT REPEAT 788 821 TPR 9.
 SQ SEQUENCE 1307 AA; 141286 MW; 6FBSASF38AD9FE53 CRC64;
 Query Match 5.7%; Score 94; DB 1; Length 1307;
 Best Local Similarity 22.5%; Pred. No. 13;
 Matches 80; Conservative 34; Mismatches 125; Indels 116; Gaps 15;
 QY 31 ILSRH---GVRSPTKALMVSPDWP---WPVGLTPRGAALVILGVRYVFARGLLPGCCPAA 83
 Db 349 LLSRMLDAAAREFHRAVDIDPHDPDALGGGLVAQARQQPALARQYFLOAQAGPDAAG 408
 QY 84 TILADVDERTRTQAFAGLAGPGCAIVHAGDDP---IFHGLDTGCLDQADAILAA----- 135
 Db 409 HWRAAL-----KALETG-----GGVDPLVARIVQAINAGRYDAARADLATLGR 453
 QY 136 -GEG-----GLTAREGLTLAKVLNFASACLECARVVG----- 166
 Db 454 POGYGLTVLSQAALARRQDGTADAVRLYREVVRAPRDAGALFSLGALDVQVGATEAAD 513
 QY 167 -----RLGPLLASTLSEIFLLEYAQGPMEVGMGRISAEMLLHNAQFLMNRTPYIA----- 216
 Db 514 ILTRQLRLAPAMARRELEAMMLSAQADRAGDD-D-GRIA---LLRRAQALDDPDVVRKLIA 569
 QY 217 -----RATPILIVTALSAPAR-----VLLAGHDTNLAGGLDMLWOLPOPDE 257
 Db 570 HALDDAGDHAAQAQAMMDALTAPRNASQAQALQAGIYAMGRHDT--ATAGALLERMPTRGR 627
 QY 258 TPGGCGALVFELWNRYRVVYQTMQDLRLNLEPLLPRILPI---PGCGSEAACSLSD 309
 Db 628 TP-----DMDRLASLVLDQRIAAALNHAPVAGNAVALAD 663

Search completed: April 30, 2004, 12:39:31
 Job time : 9.7561 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 30, 2004, 12:34:35 ; Search time 28,8084 Seconds
(without alignments)
3482.836 Million cell updates/sec

Title: US-10-021-723B-14
Perfect score: 1647
Sequence: 1 VSLLAALLAALAPAAAA.....CGSEAACSLSDFARLVAPAC 318

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- 1: SP_TREMBL 25:*
 - 2: sp_archaea:*
 - 3: sp_bacteria:*
 - 4: sp_fungi:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_nhc:*
 - 8: sp_organelle:*
 - 9: sp_phage:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*
 - 15: sp_rvirus:*
 - 16: sp_bacteriap:*
 - 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	613.5	37.2	432	2 Q8K07	Q8kd7 escherichia
2	603.5	37.0	432	2 Q8N88	Q8gn88 escherichia
3	607.5	36.9	432	2 Q8K06	Q8kx6 escherichia
4	607.5	36.9	432	2 Q8K08	Q8kx8 escherichia
5	607.5	36.9	432	2 Q8K0E	Q8kx0 escherichia
6	606.5	36.8	432	2 Q8K0D	Q8kx9 escherichia
7	606.5	36.8	434	16 Q8XC29	Q8xc29 escherichia
8	605.5	36.8	446	16 Q8CW75	Q8cw75 escherichia
9	603.5	36.6	432	2 Q8K05	Q8kx5 escherichia
10	600.5	36.5	442	16 Q8K02	Q8kx2 escherichia
11	596.5	36.2	432	16 Q7UD08	Q7ud08 shigella fl
12	590	35.8	441	16 Q8ZFP6	Q8zfp6 yersinia pe
13	377.5	22.9	414	16 Q9AAQ4	Q9aac4 caulobacter
14	359	21.8	443	16 Q8P330	Q8p330 xanthomonas
15	353	21.4	435	16 Q8P53	Q8p53 xanthomonas
16	348	21.1	392	16 Q8P76	Q8p76 xanthomonas

17	325	19.7	414	16 Q8PPX6	Q8ppx6 xanthomonas
18	290.5	17.6	413	16 Q7UD02	Q7ud02 shigella fl
19	290.5	17.6	421	16 Q83RV6	Q83rv6 shigella fl
20	289.5	17.6	413	16 Q8XB26	Q8xb26 escherichia
21	278.5	16.9	413	16 Q8Z7P1	Q8z7p1 salmonella
22	262.5	15.9	421	2 Q84CN9	Q84cn9 klebsiella
23	260.5	15.8	428	2 Q8GD20	Q8gd20 pseudomonas
24	249.5	15.1	421	2 Q7WSY1	Q7wsy1 klebsiella
25	220	13.4	318	16 Q8CW72	Q8cw72 escherichia
26	148	9.0	419	2 Q8VQS2	Q8vqs2 klebsiella
27	138	8.4	426	4 Q8BZG2	Q8bzg2 homo sapien
28	119.5	7.3	333	4 Q8BZG3	Q8bzg3 homo sapien
29	117	7.1	395	5 Q9VD68	Q9vd68 drosophila
30	117	7.1	502	16 Q91LD5	Q91ld5 pseudomonas
31	111.5	6.8	811	2 Q8KM35	Q8km35 myxococcus
32	111	6.7	1589	2 Q8KUF5	Q8kuf5 actinosyne
33	110	6.7	435	5 Q8MY06	Q8my06 drosophila
34	109.5	6.6	757	6 Q9N015	Q9n015 macaca fasc
35	107.5	6.5	426	16 Q9A937	Q9a937 caulobacter
36	107	6.5	818	16 Q9F3E4	Q9f3e4 streptomyce
37	106.5	6.5	352	2 Q9APF7	Q9apf7 legionella
38	106.5	6.5	928	16 Q92KH3	Q92kh3 rhizobium m
39	106	6.4	475	16 Q8XW24	Q8xw24 ralstonia s
40	106	6.4	972	16 Q8U5J7	Q8u5j7 agrobacteri
41	106	6.4	994	16 Q8UH19	Q8uh19 agrobacteri
42	106	6.4	1531	2 Q93N88	Q93n88 streptomyce
43	104.5	6.3	386	4 Q96QK9	Q96qk9 homo sapien
44	104	6.3	358	17 Q9YBP4	Q9ybp4 aeropyrum p
45	104	6.3	9507	2 Q9EWA1	Q9ewal streptomyce

ALIGNMENTS

RESULT 1

Q8K07 ID Q8K07 PRELIMINARY; PRT; 432 AA.
AC Q8K07; DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Periplasmic phosphoanhydride phosphohydrolase.
GN APPA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90368616; PubMed=2168385;
RA Dassa J., Marck C., Boquet P.-L.L.;
RT "The complete nucleotide sequence of the Escherichia coli gene appA
RT reveals significant homology between pH 2.5 acid phosphatase and
RT Glucose-1-phosphatase."
RL J. Bacteriol. 172:5497-5500 (1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93054596; PubMed=1429631;
RA Ostain K., Harms E.H., Stevis P.E., Kuciel R., Zhou M.M.,
RA Van Etten R.L.;
RT "Overexpression, site-directed mutagenesis, and mechanism of
RT Escherichia coli acid phosphatase."
J. Biol. Chem. 267:22830-22836 (1992).
RL EMBL; L03373; AAA0005.1; -
DR GO; GO:0003993; F:acid phosphatase activity; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR000560; HisAc_phsptase.
DR Pfam; PF00328; acid_phosphat; 1.
DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
DR Hydrolase.
SQ SEQUENCE 432 AA; 46971 MW; 755D5E4B1AD916A6 CRC64;

Query Match 37.2%; Score 613.5; DB 2; Length 432;
Best Local Similarity 45.3%; Pred. No. 9.3e-39;
Matches 194; Conservative 44; Mismatches 79; Indels 111; Gaps 36;

QY 2 SLLLAALALALAP-AAAAAP-LKLEKVVILSRHGRSPTKA--LM--VSPD-WP-W 52
DB 3 AILIPFLSLIPLTPQSAFAQSEBELKLESVVIVSRHGRVAPTKATQLMQDVTDPDAWPTW 62
QY 53 PV---GLTPRGAALV-LLGY---RYFARGLL--PGCPAAG--TILADVDERTR-TGOAF 99
DB 63 PVKLGWLTFRGELLAYLGHYQORLVADGLAKKGCQSQGVAVIADVDERTRKTGEAF 122
QY 100 AAGLAPGCAI-VH-----AGDDPIPHGLDTG-C-LDQA---DAILAAGEGGLT--ARHGL 146
DB 123 AAGLAPDCAITVHTQADTSSPDPLFNPLKTVGCQLDNANVTDAILSRAGGSIAFTGHRQ 182
QY 147 T-----LAKVLFN--ASACLE-----CARVVGRLGPI--LASTLSEI 179
DB 183 TAFRELRVLPNPNQSLCKREKQDECSLTQALPSELKVSADVSLTGAVSLASMLTEI 242
QY 180 FLLEYAQPMVEVWGRI--SAEW-----LLHNAQF-LMNRTPYIA--RATPI--LIVTALS 228
DB 243 FLQQAQMGPEPGMGRIITDSHWNLTLLSLHNAQFYLLQRTPEVARSRAATPLLDLIKALT 302
QY 229 -----PARVVLLAGHDTNLA-LGG-LDL-WQLP-OPDETPPGGALVFELWNR 271
DB 303 PHPPQKQAYGVTLPTSVLFIAGHDTNLANLGALLENWTLPGQPDNTPPGGELVFERWR 362
QY 272 -----YRVVMYQTMQDLRLNLEPLPRILP-----IPGC---GSEAACSLSDFARLV 314
DB 363 LSDNSQWIOVSLVFTLQMRDKTFLSLNTPPGEVKLTLAGCEERNAQGMCSLAGTQIV 422
QY 315 ----APAC 318
DB 423 NEARIPAC 430

RESULT 2
Q8GN88 PRELIMINARY; PRT; 432 AA.
ID Q8GN88
AC Q8GN88
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE APPA.
GN APPA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90368616; PubMed=2168385;
RA Dassa J., Marck C., Boquet P., L.L.;
RT "The complete nucleotide sequence of the Escherichia coli gene appA
reveals significant homology between pH 2.5 acid phosphatase and
glucose-1-phosphatase.";
RL J. Bacteriol. 172:5497-5500 (1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93054596; PubMed=1429631;
RA Ostanin K., Harms E.H., Stevis P.E., Kuciel R., Zhou M.M.,
RA Van Etten R.L.;
RT "Overexpression, site-directed mutagenesis, and mechanism of
Escherichia coli acid phosphatase.";
RL J. Biol. Chem. 267:22830-22836 (1992).
DR EMBL; L03374; AAC0006.1;
DR GO; GO:0003993; F:acid phosphatase activity; IEA.
DR InterPro; IPR000560; HisAc_phsphtse.
DR PROSITE; PS00616; HIS ACID PHOSPHAT 1;
DR PROSITE; PS00778; HIS ACID PHOSPHAT 2;
SQ SEQUENCE 432 AA; 47055 MW; 5B355D76E737737737 CRC64;

Query Match 37.0%; Score 609.5; DB 2; Length 432;
Best Local Similarity 45.1%; Pred. No. 1.9e-38;
Matches 193; Conservative 45; Mismatches 79; Indels 111; Gaps 36;

QY 2 SLLLAALALALAP-AAAAAP-LKLEKVVILSRHGRSPTKA--LM--VSPD-WP-W 52
DB 3 AILIPFLSLIPLTPQSAFAQSEBELKLESVVIVSRHGRVAPTKATQLMQDVTDPDAWPTW 62
QY 53 PV---GLTPRGAALV-LLGY---RYFARGLL--PGCPAAG--TILADVDERTR-TGOAF 99
DB 63 PVKLGWLTFRGELLAYLGHYQORLVADGLAKKGCQSQGVAVIADVDERTRKTGEAF 122
QY 100 AAGLAPGCAI-VH-----AGDDPIPHGLDTG-C-LDQA---DAILAAGEGGLT--ARHGL 146
DB 123 AAGLAPDCAITVHTQADTSSPDPLFNPLKTVGCQLDNANVTDAILSRAGGSIAFTGHRQ 182
QY 147 T-----LAKVLFN--ASACLE-----CARVVGRLGPI--LASTLSEI 179
DB 183 TAFRELRVLPNPNQSLCKREKQDECSLTQALPSELKVSADVSLTGAVSLASMLTEI 242
QY 180 FLLEYAQPMVEVWGRI--SAEW-----LLHNAQF-LMNRTPYIA--RATPI--LIVTALS 228
DB 243 FLQQAQMGPEPGMGRIITDSHWNLTLLSLHNAQFYLLQRTPEVARSRAATPLLDLIKALT 302
QY 229 -----PARVVLLAGHDTNLA-LGG-LDL-WQLP-OPDETPPGGALVFELWNR 271
DB 303 PHPPQKQAYGVTLPTSVLFIAGHDTNLANLGALLENWTLPGQPDNTPPGGELVFERWR 362
QY 272 -----YRVVMYQTMQDLRLNLEPLPRILP-----IPGC---GSEAACSLSDFARLV 314
DB 363 LSDNSQWIOVSLVFTLQMRDKTFLSLNTPPGEVKLTLAGCEERNAQGMCSLAGTQIV 422
QY 315 ----APAC 318
DB 423 NEARIPAC 430

RESULT 3
Q8RKD6 PRELIMINARY; PRT; 432 AA.
ID Q8RKD6
AC Q8RKD6
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Periplasmic phosphoanhydride phosphohydrolase.
GN APPA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90368616; PubMed=2168385;
RA Dassa J., Marck C., Boquet P., L.L.;
RT "The complete nucleotide sequence of the Escherichia coli gene appA
reveals significant homology between pH 2.5 acid phosphatase and
glucose-1-phosphatase.";
RL J. Bacteriol. 172:5497-5500 (1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93054596; PubMed=1429631;
RA Ostanin K., Harms E.H., Stevis P.E., Kuciel R., Zhou M.M.,
RA Van Etten R.L.;
RT "Overexpression, site-directed mutagenesis, and mechanism of
Escherichia coli acid phosphatase.";
RL J. Biol. Chem. 267:22830-22836 (1992).
DR EMBL; L03374; AAC0006.1;
DR GO; GO:0003993; F:acid phosphatase activity; IEA.
DR InterPro; IPR000560; HisAc_phsphtse.
DR PROSITE; PS00616; HIS ACID PHOSPHAT 1;
DR PROSITE; PS00778; HIS ACID PHOSPHAT 2;
KW Hydrolase.
SQ SEQUENCE 432 AA; 46971 MW; AF86C41EA6193AC5 CRC64;

Query Match 36.9%; Score 607.5; DB 2; Length 432;
 Best Local Similarity 45.1%; Pred. No. 2.7e-38;
 Matches 193; Conservative 44; Mismatches 80; Indels 111; Gaps 36;

QY 2 SLLAALLAALAP-AAAAAP-LKLEKVVILSRHGVSPKTA--LM--VSPD-WP-W 52
 DB 3 AILIPFLSLIPLTQSAFAQSEPELKLESVVIVSRHGVAPTKATQMLQDVTDPDAPTW 62
 QY 53 PV---GLTPRGAALV-LIGY---RYFARGLL--PGCPAAG--TILADVDETR-TQAP 99
 DB 63 PVKLGWLTFRGELIAYLGHYQORLVADGLAKKGCPOSGQVAVIADVDETRTKTGEAF 122
 QY 100 AAGLAPGCAI-VH-----AGDDPIPHGLDTG-C-LDOA---DAILAAGEGGLT--ARHGL 146
 DB 123 AAGLAPDCAITVHTQADTSPPDLFNLKTVGCVQCLDNANVTDAILSRAGSIADFTGHRQ 182
 QY 147 T----LAKYVNF--ASACLE-----CARVVGRIGLPL-LASTLSI 179
 DB 183 TAFRELERVLPFQSNLCKREKQDESCSLTQALPSELKVSADNVSLTGAVSLASMLTEI 242
 QY 180 FLLEYAOGPMVGVGRI--SAEW---LLHNAQF-LMNRTPVIA--RATPI--LIVTALS 228
 DB 243 FLIQAQGMPEPGNGRIITDHSQWNTLLSLHNAQFYLLQRTPEVARSRATPLDLIKTALT 302
 QY 229 -----PARVLLAGHTNLA-LGG-LDL-WQLP-QDETTPPGALVFELWNR 271
 DB 303 PHPPQKQAYGVTLPSTVLFIAGHDTNLANLGALSLNWTLPQPDNTPPGSELVFERWR 362
 QY 272 -----YRVVMYQTMQDLNLEPLPRILP-----IPGC---GSEAACSLSDFARLV 314
 DB 363 LSDNSQWQVSLVFTLQMDKTPSLNTPPGEVKLTLAGCEBNAQGMCSLAGFTQIV 422
 QY 315 ----APAC 318
 DB 423 NEARIPAC 430

RESULT 4

QSRK08 PRELIMINARY; PRT; 432 AA.
 AC QSRK08;
 DT 01-JUN-2002 (TREMELrel. 21, Created)
 DT 01-JUN-2002 (TREMELrel. 21, Last sequence update)
 DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
 DE Periplasmic phosphoanhydride phosphohydrolase.
 GN APPA.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90368616; PubMed=2168385;
 RA Dassa J., Marck C., Boquet P.-L.L.;
 RT "The complete nucleotide sequence of the Escherichia coli gene appA
 RT reveals significant homology between pH 2.5 acid phosphatase and
 RT glucose-1-phosphatase."
 RL J. Bacteriol. 172:5497-5500(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93054596; PubMed=1429631;
 RA Ostanin K., Harms E.H., Stevis P.E., Kuciel R., Zhou M.M.,
 RT "Overexpression, site-directed mutagenesis, and mechanism of
 RT Escherichia coli acid phosphatase."
 RL J. Biol. Chem. 267:22830-22836(1992).
 DR EMBL; L03372; AAA00004.1; -
 DR GO; GO:0003993; F:acid phosphatase activity; IEA.
 DR GO; GO:0016787; F:hydrolase activity; IEA.
 DR InterPro; IPR00560; HisAc_phsphtase.
 DR Pfam; PF00328; acid_phosphat_1.
 DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
 KW Hydrolase.

SQ SEQUENCE 432 AA; 46971 MW; 9A85536B57FCCFB5 CRC64;
 Query Match 36.9%; Score 607.5; DB 2; Length 432;
 Best Local Similarity 45.1%; Pred. No. 2.7e-38;
 Matches 193; Conservative 44; Mismatches 80; Indels 111; Gaps 36;

QY 2 SLLAALLAALAP-AAAAAP-LKLEKVVILSRHGVSPKTA--LM--VSPD-WP-W 52
 DB 3 AILIPFLSLIPLTQSAFAQSEPELKLESVVIVSRHGVAPTKATQMLQDVTDPDAPTW 62
 QY 53 PV---GLTPRGAALV-LIGY---RYFARGLL--PGCPAAG--TILADVDETR-TQAP 99
 DB 63 PVKLGWLTFRGELIAYLGHYQORLVADGLAKKGCPOSGQVAVIADVDETRTKTGEAF 122
 QY 100 AAGLAPGCAI-VH-----AGDDPIPHGLDTG-C-LDOA---DAILAAGEGGLT--ARHGL 146
 DB 123 AAGLAPDCAITVHTQADTSPPDLFNLKTVGCVQCLDNANVTDAILSRAGSIADFTGHRQ 182
 QY 147 T----LAKYVNF--ASACLE-----CARVVGRIGLPL-LASTLSI 179
 DB 183 TAFRELERVLPFQSNLCKREKQDESCSLTQALPSELKVSADNVSLTGAVSLASMLTEI 242
 QY 180 FLLEYAOGPMVGVGRI--SAEW---LLHNAQF-LMNRTPVIA--RATPI--LIVTALS 228
 DB 243 FLIQAQGMPEPGNGRIITDHSQWNTLLSLHNAQFYLLQRTPEVARSRATPLDLIKTALT 302
 QY 229 -----PARVLLAGHTNLA-LGG-LDL-WQLP-QDETTPPGALVFELWNR 271
 DB 303 PHPPQKQAYGVTLPSTVLFIAGHDTNLANLGALSLNWTLPQPDNTPPGSELVFERWR 362
 QY 272 -----YRVVMYQTMQDLNLEPLPRILP-----IPGC---GSEAACSLSDFARLV 314
 DB 363 LSDNSQWQVSLVFTLQMDKTPSLNTPPGEVKLTLAGCEBNAQGMCSLAGFTQIV 422
 QY 315 ----APAC 318
 DB 423 NEARIPAC 430

RESULT 5

QSRK08 PRELIMINARY; PRT; 432 AA.
 AC QSRK08;
 DT 01-JUN-2002 (TREMELrel. 21, Created)
 DT 01-JUN-2002 (TREMELrel. 21, Last sequence update)
 DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
 DE Periplasmic phosphoanhydride phosphohydrolase.
 GN APPA.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90368616; PubMed=2168385;
 RA Dassa J., Marck C., Boquet P.-L.L.;
 RT "The complete nucleotide sequence of the Escherichia coli gene appA
 RT reveals significant homology between pH 2.5 acid phosphatase and
 RT glucose-1-phosphatase."
 RL J. Bacteriol. 172:5497-5500(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93054596; PubMed=1429631;
 RA Ostanin K., Harms E.H., Stevis P.E., Kuciel R., Zhou M.M.,
 RT "Overexpression, site-directed mutagenesis, and mechanism of
 RT Escherichia coli acid phosphatase."
 RL J. Biol. Chem. 267:22830-22836(1992).
 DR EMBL; L03370; AAA00002.1; -
 DR GO; GO:0003993; F:acid phosphatase activity; IEA.
 DR GO; GO:0016787; F:hydrolase activity; IEA.
 DR InterPro; IPR00560; HisAc_phsphtase.
 DR Pfam; PF00328; acid_phosphat_1.
 DR

DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
KW Hydrolase. 432 AA; 46971 MW; 5BBD632D4682EABF CRC64;
SQ SEQUENCE

Query Match 36.9%; Score 607.5; DB 2; Length 432;
Best Local Similarity 45.1%; Pred. No. 2.7e-38;
Matches 193; Conservative 44; Mismatches 80; Indels 111; Gaps 36;

QY 2 SLILALLAALAP-AAAAAP-LKLEKWLHSGVSPKA--LM--VSPD-WP-W 52
DB 3 AILPFLSLIPIPTQSAFAQSEPELKESVIVSANGVRAPTKATQLMQDTPDAPTW 62
QY 53 PV---GLTPRGAALV-LLGY---RYFARGLL--PGCPAAG--TILADVDETR-TGQAF 99
DB 63 PVKLGWLTPRGGELIAYLGHYQRLVADGLLAKGCPQSGQVAILADVDETRTKTGEAF 122
QY 100 AAGLAPGCAI-VH-----AGDDPIFHGLDTG-C-LDOA---DAILAAGEGLT--ARHGL 146
DB 123 AAGLAPGCAITVHTQADTSPPDLFNLKTVGVCLDNNVNTAILSRAGGSTADTGHQ 182
QY 147 T-----LAKVNF--ASACLE-----CARVVGRLGFL--LASTLSEI 179
DB 183 TAPRELVNFPQSNLCKREKQDESCSLTQALPSELKVSADNVSLTGAVSLASMLTEI 242
QY 180 FLLEYAQGPMEVGVGRI--SAEW---LLHNAQF-LMNRTPYA--RATPI--LIVTALS 228
DB 243 FLQQAQGMPEPGWGRITDSHQNTLLSHNAQFYLLQRTPEVARSRAPLDDLIKALT 302
QY 229 -----PARVVLLAGHDTNLA-LGG-LDL-WQLP-QPDETPPGGALVPELWNR 271
DB 303 PHPPQQAQVGLTPTSVLFTAGHDTNLANLGGALELNTWLPFGQPDNTPPGGELVFERWR 362
QY 272 -----YRVVMYQMDQLNLEPLPRILP-----IPGC---GSEAACSLSDFARLV 314
DB 363 LSDNSQWISVFTLQQRDKTPLSLNTPPGVEVKTLGACERNAQGMCSLAGTQIV 422
QY 315 ----APAC 318
DB 423 NEARIPAC 430

RESULT 6
Q8KXD9 PRELIMINARY; PRT; 432 AA.
ID Q8KXD9
AC Q8KXD9
DT 01-JUN-2002 (T-EMBLrel. 21, Created)
DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Periplasmic phosphoanhydride phosphohydrolase.
GN APPA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90368616; PubMed=2168385;
RA Dassa J., Marck C., Boquet P.-L.L.;
RT "The complete nucleotide sequence of the Escherichia coli gene appa
RT reveals significant homology between pH 2.5 acid phosphatase and
RT glucose-1-phosphatase".
RL J. Bacteriol. 172:5497-5500 (1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90354596; PubMed=1429631;
RA Ostalan K., Harms E.H., Stevis P.E., Kuciel R., Zhou M.M.,
RA Van Etten R.L.;
RT "Overexpression, site-directed mutagenesis, and mechanism of
RT Escherichia coli acid phosphatase".
RL J. Biol. Chem. 267:22830-22836 (1992).
DR EMBL; L03371; AAA00003.1; -
DR GO; GO:0003993; F:acid phosphatase activity; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.

DR InterPro; IPR000560; HisAc phsphtse.
DR Pfam; PF00328; acid_phosphat; 1.
KW PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
KW Hydrolase.
SQ SEQUENCE

Query Match 36.8%; Score 606.5; DB 2; Length 432;
Best Local Similarity 45.1%; Pred. No. 3.2e-38;
Matches 193; Conservative 45; Mismatches 79; Indels 111; Gaps 36;

QY 2 SLILALLAALAP-AAAAAP-LKLEKWLHSGVSPKA--LM--VSPD-WP-W 52
DB 3 AILPFLSLIPIPTQSAFAQSEPELKESVIVSANGVRAPTKATQLMQDTPDAPTW 62
QY 53 PV---GLTPRGAALV-LLGY---RYFARGLL--PGCPAAG--TILADVDETR-TGQAF 99
DB 63 PVKLGWLTPRGGELIAYLGHYQRLVADGLLAKGCPQSGQVAILADVDETRTKTGEAF 122
QY 100 AAGLAPGCAI-VH-----AGDDPIFHGLDTG-C-LDOA---DAILAAGEGLT--ARHGL 146
DB 123 AAGLAPGCAITVHTQADTSPPDLFNLKTVGVCLDNNVNTAILSRAGGSTADTGHQ 182
QY 147 T-----LAKVNF--ASACLE-----CARVVGRLGFL--LASTLSEI 179
DB 183 TAPRELVNFPQSNLCKREKQDESCSLTQALPSELKVSADNVSLTGAVSLASMLTEI 242
QY 180 FLLEYAQGPMEVGVGRI--SAEW---LLHNAQF-LMNRTPYA--RATPI--LIVTALS 228
DB 243 FLQQAQGMPEPGWGRITDSHQNTLLSHNAQFYLLQRTPEVARSRAPLDDLIKALT 302
QY 229 -----PARVVLLAGHDTNLA-LGG-LDL-WQLP-QPDETPPGGALVPELWNR 271
DB 303 PHPPQQAQVGLTPTSVLFTAGHDTNLANLGGALELNTWLPFGQPDNTPPGGELVFERWR 362
QY 272 -----YRVVMYQMDQLNLEPLPRILP-----IPGC---GSEAACSLSDFARLV 314
DB 363 LSDNSQWISVFTLQQRDKTPLSLNTPPGVEVKTLGACERNAQGMCSLAGTQIV 422
QY 315 ----APAC 318
DB 423 NEARIPAC 430

RESULT 7
Q8KXD9 PRELIMINARY; PRT; 434 AA.
ID Q8KXD9
AC Q8KXD9
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Phosphoanhydride phosphorylase, pH 2.5 acid phosphatase,
DE periplasmic.
GN APPA OR Zi397 OR ECS1136.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Rose D.J., Mayhew G.F., Klink S., Boutin A., Shao Y., Miller L.,
RA Postel G., Hackett J., Davis N.W., Lim A., Dimalanta E.T., Potamocis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RL Nature 409:529-533 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=O157:H7 / RIMD 0509952;
RC MEDLINE=21156231; PubMed=11258796;

RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tohe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kunara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic *Escherichia coli*
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
DR EMBL; AE005292; AAG55528.1; ALT_INIT.
DR EMBL; AP002554; BAB34559.1; -.
DR PIR; D85633; D85633.
DR PIR; H90770; H90770.
DR GO; GO:0003993; F:acid phosphatase activity; IEA.
DR InterPro; IPR000560; HisAcidPhosphatase.
DR Pfam; PF00328; acid phosphatase; 1.
DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
KW Complete proteome.
SQ SEQUENCE 434 AA; 47337 MW; F197DF7D1869F9C4 CRC64;

Query Match 36.8%; Score 606.5; DB 16; Length 434;
Best Local Similarity 44.9%; Pred. No. 3.2e-38;
Matches 193; Conservative 49; Mismatches 81; Indels 113; Gaps 36;

QY 2 SLLLAALLAALAPAAAAA-----EP-LKLEKVVILSRHGVRSPTKA--LM--VSPD-WP 51
DB 3 AILIPFLSLLIPLTPQSAFAQSEPELKLESVVIVSRHGVRAPTKATQLMQDVTDPDAMP 62
QY 52 -WPV---GLTPRGAALV-LIGY-----RYFARGLL--PGCPAAG--TILADVDERTR-TQ 97
DB 63 NMPVKGWLTFRGGLIAYLGHYQRLVADGLLTKKGCPOQGVAILADVDERTRKTGE 122
QY 98 AFAAGLAPGCAI-VH-----AGDDPIFGHLDTG-C-LDQA---DAILAAGEGLT--ARH 144
DB 123 AFAAGLAPDCAITVHTQADTSPPDPLFNPLKTVGCQDANVNTDAILSRAGGSIADFTGH 182
QY 145 GLT-----LAKVLFN--ASACLE-----RYFARGLL--PGCPAAG--TILADVDERTR-TQ 177
DB 183 RQAFRELERLVNFPQSNCLNREKODESCSLTQALPSELKVSADNVSLTGAVSLASMLT 242
QY 178 EIFLEVAQPMVEGWGRI--SAEW-----LHNAQF-LMNRTPYIA--RATPI--LIIVA 226
DB 243 EIFLLOAQGMPEGWGRIITDSHWNLTLLSHNAQFVLLQRTPEVARSRATPLDLIMIA 302
QY 227 LS-----PARVVLGAGHTNLA-LGG-LDL-WQLP-OPDETTPPGALVFEW 269
DB 303 LTHPPKQAYGVTLPTSVLFIAGHTNLANLGALLENLTLPGQDNTTPPGELVFERW 362
QY 270 NR-----YVRVYQTMQDLNLEPLLPRIPL-----IPGC---GSEAAACSLDFA 312
DB 363 RRLSDNSQWISVLFQTLQOMRDKTFLSLNTPGVEVKLTLAGCEERNAQCMCSLAGFTQ 422
QY 313 LV-----APAC 318
DB 423 IVNEARIPAC 432

RESULT 8
Q8CW75 PRELIMINARY; PRT; 446 AA.
AC Q8CW75;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Periplasmic appA protein precursor.
GN APPA OR C1121.
OS *Escherichia coli* O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;

Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Raeko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic *Escherichia coli*."
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AE016758; AAN79589.1; -.
DR GO; GO:0003993; F:acid phosphatase activity; IEA.
DR InterPro; IPR000560; HisAcidPhosphatase.
DR Pfam; PF00328; acid phosphatase; 1.
DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
KW Complete proteome.
SQ SEQUENCE 446 AA; 48605 MW; F1308CD691DB9F49 CRC64;

Query Match 36.8%; Score 605.5; DB 16; Length 446;
Best Local Similarity 44.9%; Pred. No. 3.9e-38;
Matches 192; Conservative 44; Mismatches 81; Indels 111; Gaps 36;

QY 2 SLLLAALLAALAPAAAAA-----EP-LKLEKVVILSRHGVRSPTKA--LM--VSPD-WP-W 52
DB 13 AILIPFLSLLIPLTPKSAFAQSEPELKLESVVIVSRHGVRAPTKATQLMQDVTDPDAMP 72
QY 53 PV---GLTPRGAALV-LIGY-----RYFARGLL--PGCPAAG--TILADVDERTR-TQ 99
DB 73 PVKLGWLTFRGGLIAYLGHYQRLVADGLLTKKGCPOQGVAILADVDERTRKTGEAF 132
QY 100 AAGLAPGCAI-VH-----AGDDPIFGHLDTG-C-LDQA---DAILAAGEGLT--ARHGL 146
DB 133 AAGLAPDCAITVHTQADTSPPDPLFNPLKTVGCQDANVNTDAILSRAGGSIADFTGHRQ 192
QY 147 T-----LAKVLFN--ASACLE-----CARVVGRLGPL-LASTLSEI 179
DB 193 TAFRELERLVNFPQSNCLNREKODESCSLTQALPSELKVSADNVSLTGAVSLASMLT 252
QY 180 EIFLEVAQPMVEGWGRI--SAEW-----LHNAQF-LMNRTPYIA--RATPI--LIIVA 228
DB 253 EIFLLOAQGMPEGWGRIITDSHWNLTLLSHNAQFVLLQRTPEVARSRATPLDLIMAALT 312
QY 229 LS-----PARVVLGAGHTNLA-LGG-LDL-WQLP-OPDETTPPGALVFEW 271
DB 313 PHPSKQAYGVTLPTSVLFIAGHTNLANLGALLENLTLPGQDNTTPPGELVFERWR 372
QY 272 NR-----YVRVYQTMQDLNLEPLLPRIPL-----IPGC---GSEAAACSLDFA 314
DB 373 LSDNSQWISVLFQTLQOMRDKTFLSLNTPGVEVKLTLAGCEERNAQCMCSLAGFTQIV 432
QY 315 LV-----APAC 318
DB 433 IVNEARIPAC 440

RESULT 9
Q8RKD5 PRELIMINARY; PRT; 432 AA.
AC Q8RKD5;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Periplasmic phosphoanhydride phosphohydrolase.
GN APPA.
OS *Escherichia coli*.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90368616; PubMed=2168385;
RA Passa J., Marck C., Boquet P.-L.;
RT "The complete nucleotide sequence of the *Escherichia coli* gene appA
RT reveals significant homology between pH 2.5 acid phosphatase and
RT glucose-1-phosphatase.";

J. Bacteriol. 172:5497-5500 (1990).

[2]

SEQUENCE FROM N.A.

MEDLINE=93054596; PubMed=1429631;

Ostain K., Harms E.H., Stevis P.E., Zhou M.M.,

Van Etten R.L.;

"Overexpression, site-directed mutagenesis, and mechanism of

Escherichia coli acid phosphatase";

J. Biol. Chem. 267:22830-22836 (1992).

ENBL; L03375; AAA00007.1; -

DR GO: 0003993; F:acid phosphatase activity; IEA.

DR GO: 0016787; F:hydrolase activity; IEA.

DR InterPro: IPR000560; HisAc_phsphtse.

DR Pfam: PF00328; acid phosphat; 1.

DR PROSITE; PS00616; acid phosphat; 1.

KW Hydrolase.

SEQUENCE. 432 AA; 46990 MW; 951F393EA9A1A47C CRC64;

Query Match 36.6%; Score 603.5; DB 2; Length 432;

Best Local Similarity 45.1%; Pred. No. 5.4e-38;

Matches 193; Conservative 44; Mismatches 80; Indels 111; Gaps 36;

QY 2 SLLLAALLAALAP--AAAAAEP-LKLEKVVILSRHGVSPTKA--LM--VSPD-WP-W 52

Db 3 AILIPFLSLIPLTPQSAFAQSEPELKLESVVIVSRHGVRAFTKATQLMODVTDAMPTW 62

QY 53 PV---GLTPRGALV-LIGY----RYFAGLL--FGCPAAG--TILADVDERTR-TQAF 99

Db 63 PVKGLWLTFRGELIAYLGHYQORLVADGLLAKKGCPSQGVAVIADVDERTRKTGEAF 122

QY 100 AAGLAPGCAI-VH-----AGDDPIFGHLDTC-C-LDQA---DAILAAGEGLT--ARHGL 146

Db 123 AAGLAPDCAITVHTQADTSSPDPLFNPLKTVGCQLDNANVTDAILSRAGSGIAFTGHQ 182

QY 147 T---LAKVNF--ASACLE-----CARVVGRLGPI--LASTLSEI 179

Db 183 TAFRELVNFPQSNLCNREKQDECSLTQALPSELKVSADNVSLTGAVSLASMLTEI 242

QY 180 FILEYAQGMVEVWGRI--SAEW---LLHNAQF-LMNRTPYIA--RATPI--LIVTALS 228

Db 243 FLIQAQGMPEPGWGRI--TDSHWNLTLSLHNAQFYLQRTPEVARSRAATPLLDLIKALT 302

QY 229 -----PARVVLLAGHDNLA-LGG-LDL-WQLP-QDETTPPGALVPELWNR 271

Db 303 PHPPQKQAVGVTLPSTVLFIAAGDNLNLAGLELWNLTPGQDNTPPGELVFERWR 362

QY 272 -----YVRVYQTMQDLRNLPLPRILP-----IPGC---GSEAACSLSDPARLV 314

Db 363 LSDNSQWIVQSLVFTLQQRMDKTPSLNTPPGVVKLTLAGCEERNAQGMCSLAGFTQIV 422

QY 315 ----APAC 318

Db 423 NEARIPAC 430

RESULT 10

Q83RW2 PRELIMINARY; PRT; 442 AA.

ID Q83RW2

AC Q83RW2; (TREMblrel. 24, Created)

DT 01-JUN-2003 (TREMblrel. 24, Last sequence update)

DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)

DE Phosphoanhydride phosphorylase, pH 2.5 acid phosphatase,

DE periplasmic.

OS *Shigella flexneri*.

GN APPA OR SF0982.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Shigella.

OX NCBI_TaxID=623;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=301 / Serotype 2a;

RC MEDLINE=22272406; PubMed=12384590;

RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,

RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,

RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,

RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,

RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,

RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,

RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,

RA Yu J.,

"Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity

through comparison with genomes of *Escherichia coli* K12 and O157";

RL Nucleic Acids Res. 30:4432-4441 (2002).

DR ENBL; AF015127; AAN2610.1; -

DR GO: 0003993; F:acid phosphatase activity; IEA.

DR InterPro: IPR000560; HisAc_phsphtse.

DR Pfam: PF00328; acid phosphat; 1.

DR PROSITE; PS00616; His ACID PHOSPHAT 1; 1.

DR PROSITE; PS00778; His ACID PHOSPHAT 2; 1.

KW Complete proteome.

SEQUENCE. 442 AA; 48183 MW; 20BC2C2454254680 CRC64;

Query Match 36.5%; Score 600.5; DB 16; Length 442;

Best Local Similarity 44.9%; Pred. No. 9.4e-36;

Matches 192; Conservative 43; Mismatches 82; Indels 111; Gaps 36;

QY 2 SLLLAALLAALAP--AAAAAEP-LKLEKVVILSRHGVSPTKA--LM--VSPD-WP-W 52

Db 13 AILIPFLSLIPLTPQSAFAQSEPELKLESVVIVSRHGVRAFTKATQLMODVTDAMPTW 72

QY 53 PV---GLTPRGALV-LIGY----RYFAGLL--FGCP--AAGTILADVDERTR-TQAF 99

Db 73 PVKGLWLTFRGELIAYLGHYQORLVADGLLAKKGCPSQGVAVIADVDERTRKTGEAF 132

QY 100 AAGLAPGCAI-VH-----AGDDPIFGHLDTC-C-LDQA---DAILAAGEGLT--ARHGL 146

Db 133 AAGLAPDCAITVHTQADTSSPDPLFNPLKTVGCQLDNANVTDAILSRAGSGIAFTGHQ 192

QY 147 T---LAKVNF--ASACLE-----CARVVGRLGPI--LASTLSEI 179

Db 193 TAFRELVNFPQSNLCNREKQDECSLTQALPSELKVSADNVSLTGAVSLASMLTEI 252

QY 180 FILEYAQGMVEVWGRI--SAEW---LLHNAQF-LMNRTPYIA--RATPI--LIVTALS 228

Db 253 FLIQAQGMPEPGWGRI--TDSHWNLTLSLHNAQFYLQRTPEVARSRAATPLLDLIKALT 312

QY 229 -----PARVVLLAGHDNLA-LGG-LDL-WQLP-QDETTPPGALVPELWNR 271

Db 313 PHPPQKQAVGVTLPSTVLFIAAGDNLNLAGLELWNLTPGQDNTPPGELVFERWR 372

QY 272 -----YVRVYQTMQDLRNLPLPRILP-----IPGC---GSEAACSLSDPARLV 314

Db 373 LSDNSQWIVQSLVFTLQQRMDKTPSLNTPPGVVKLTLAGCEERNAQGMCSLAGFTQIV 432

QY 315 ----APAC 318

Db 433 NEARIPAC 440

RESULT 11

Q7UD08 PRELIMINARY; PRT; 432 AA.

ID Q7UD08

AC Q7UD08; (TREMblrel. 25, Created)

DT 01-OCT-2003 (TREMblrel. 25, Last sequence update)

DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)

DE Phosphoanhydride phosphorylase, pH 2.5 acid phosphatase.

GN APPA OR S1048.

OS *Shigella flexneri*.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Shigella.

OX NCBI_TaxID=623;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=2457T / ATCC 700930 / Serotype 2a;

RC MEDLINE=22590274; PubMed=12704152;

RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,

RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,

RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,

RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,

RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,

RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,

RA Yu J.,

"Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity

through comparison with genomes of *Escherichia coli* K12 and O157";

RL Nucleic Acids Res. 30:4432-4441 (2002).

DR ENBL; AF015127; AAN2610.1; -

DR GO: 0003993; F:acid phosphatase activity; IEA.

DR InterPro: IPR000560; HisAc_phsphtse.

DR Pfam: PF00328; acid phosphat; 1.

DR PROSITE; PS00616; His ACID PHOSPHAT 1; 1.

DR PROSITE; PS00778; His ACID PHOSPHAT 2; 1.

KW Complete proteome.

SEQUENCE. 442 AA; 48183 MW; 20BC2C2454254680 CRC64;

Query Match 36.5%; Score 600.5; DB 16; Length 442;

Best Local Similarity 44.9%; Pred. No. 9.4e-36;

Matches 192; Conservative 43; Mismatches 82; Indels 111; Gaps 36;

QY 2 SLLLAALLAALAP--AAAAAEP-LKLEKVVILSRHGVSPTKA--LM--VSPD-WP-W 52

Db 13 AILIPFLSLIPLTPQSAFAQSEPELKLESVVIVSRHGVRAFTKATQLMODVTDAMPTW 72

QY 53 PV---GLTPRGALV-LIGY----RYFAGLL--FGCP--AAGTILADVDERTR-TQAF 99

Db 73 PVKGLWLTFRGELIAYLGHYQORLVADGLLAKKGCPSQGVAVIADVDERTRKTGEAF 132

QY 100 AAGLAPGCAI-VH-----AGDDPIFGHLDTC-C-LDQA---DAILAAGEGLT--ARHGL 146

Db 133 AAGLAPDCAITVHTQADTSSPDPLFNPLKTVGCQLDNANVTDAILSRAGSGIAFTGHQ 192

QY 147 T---LAKVNF--ASACLE-----CARVVGRLGPI--LASTLSEI 179

Db 193 TAFRELVNFPQSNLCNREKQDECSLTQALPSELKVSADNVSLTGAVSLASMLTEI 252

QY 180 FILEYAQGMVEVWGRI--SAEW---LLHNAQF-LMNRTPYIA--RATPI--LIVTALS 228

Db 253 FLIQAQGMPEPGWGRI--TDSHWNLTLSLHNAQFYLQRTPEVARSRAATPLLDLIKALT 312

QY 229 -----PARVVLLAGHDNLA-LGG-LDL-WQLP-QDETTPPGALVPELWNR 271

Db 313 PHPPQKQAVGVTLPSTVLFIAAG

RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Eisen J., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Nieman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eise J., Heideberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwin M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Kouri H., Shetty J., Berry K.,
RA Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.,
RT "Complete genome sequence of *Caulobacter crescentus*."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL: AE005727; AA22529.1; --
DR PIR: E87316; E87316.
DR HSP: P07102; IDKL.
DR TIGR: CC0542; --
DR GO: GO:0003993; F:acid phosphatase activity; IEA.
DR GO: GO:0016787; F:hydrolase activity; IEA.
DR InterPro: IPR000560; HisAc phosphatase.
DR Pfam: PF00328; acid phosphatase; 1.
DR PROSITE: PS00616; HIS ACID PHOSPHAT_1; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 414 AA; 42479 MW; 7105C84EA01C4051 CRC64;

Query Match 22.9%; Score 377.5; DB 16; Length 414;
Best Local Similarity 35.9%; Pred. No. 8.9e-21;
Matches 148; Conservative 47; Mismatches 112; Indels 105; Gaps 29;

QY 4 LLAALLALALAPAAAAPLKEKVVLSRHGVSPTKALMVSPD-----WP--- 51
DB 5 LRAVAVTALASAGAAAE--TLKVVLSRHGVS-----AMSPERLEASARPAPRE 58

QY 52 WPGV-LTPRGAALVL-LG--YR--YFARGLL-PCCPAAGTILADVDEPT-RTQQAFAAGL 103
DB 59 VPAGHLTARGTTLVARVGDYVRRHYAAGLLKPGDCASVYAVANVTQRTIATKAYRETL 118

QY 104 APCA-----VHAGD-PPHGLDTCGLDQADAILA-----AGE--GGLTARHGLTLAKVL 152
DB 119 APGCPVTNTVGEIGNIDMFEPVKGIV-KADHALAPAAVAGRVGGDLTAWASHNQEA 177

QY 153 NFASACLECAR-----VVGRLGP-LLASTLSEIFLLEYAOG 187
DB 178 QLDALLMQCKPCPPAPKRRVFDKPGFVGDGELAGLSPEAFASGVTESLLMAWADG 237

QY 188 P--MEVNGRISAE-----WLLHNAQF-LNRTPIYAR-----ATPILIVTALS 228
DB 238 RDFAGLWKSDEALTRSFTHQAFDLRLTPYVARTLAGHLADRLAATLRDGAAG 297

QY 229 P--ARVLLAGHDNTLA-LGGL--DLWOLF--QPDETPPGGALVFELWN-----RYVRVM 276
DB 298 PVDARLVIIAGHDGTLASGLLWELTLPQYQNPQPGGALVFERWRDDGVRVVR 357

QY 277 Y--QTMQLRNLPLPRILP-----IPGCGSAA-----CELSDFARLVAPA 317
DB 358 FTGQSLSQLRNNTALDAKTPPLSAPVFVQCGGTATPAFCRLEDFTVVRGA 409

RESULT 14
ID Q8PF30 PRELIMINARY; PRT; 443 AA.
AC Q8PF30;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Phosphoanhydride phosphohydrolase.
GN APPA OR XC00160.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=22022145; PubMed=12024217;

da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furian L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Cavarotte G., Cannavan F., Cardozo J., Chambergo F., Chapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Fexiro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.P.,
RA Locall E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C.C., Kitajima J.P.,
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463(2002).
DR EMBL: AE012175; AA040075.1; --
DR GO: GO:0003993; F:acid phosphatase activity; IEA.
DR GO: GO:0016787; F:hydrolase activity; IEA.
DR InterPro: IPR000560; HisAc phosphatase.
DR Pfam: PF00328; acid phosphatase; 1.
DR PROSITE: PS00616; HIS ACID PHOSPHAT_1; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 443 AA; 46618 MW; 8C514434CC9587CE CRC64;

Query Match 21.8%; Score 359; DB 16; Length 443;
Best Local Similarity 33.8%; Pred. No. 2.5e-19;
Matches 135; Conservative 50; Mismatches 123; Indels 92; Gaps 27;

QY 1 VSLLLAALLAAL---APAAAAPLKEKVVLSRHGVSPTKALMVSPD-----PD 49
DB 38 LGVLLVAALSIGCTPRPTPAASAS-ATLERVVVVRHGVRAPIQGEAAAAHYADAPWQ 96

QY 50 KWPVGL--TPRGAALVLLGYR---FARGLLP--GCPAAG--TILADVDEPT-RTQQA 98
DB 97 WSTPASLTVPGRKAVQSLGELVQLWQLMQALLPRNGCPARGSLTWTANTQRTIDSCAL 156

QY 99 FAAGLAPCAIVHAG-----DDPIFHGLDTCGLD-QADAILA-----GEGGLTARHG 145
DB 157 LADALAPGN-VQAGHRAVGSNDPLFRPIEAGVFPDANAATAVAATQRTGTGPAALQGHG 215

QY 146 L---TLAKVLPASACLECARVVGRL-----GPL-LASTLSEIFLLEYAOG-P 188
DB 216 AELQALQQLGCTKTPCFARMPSTLKFSADGRGLAGLPIDLASGTGTEVLLQYAEQMP 275

QY 189 M--EVNGRISAEWL-----LHNAQFLMNRTPY-----IARATPILIVTALS 233
DB 276 LQVCGWRATPERLAQVSRHLALLFDIYAREPHYMASRSCAPLAREVLQRFQDAGAP-KVS 334

QY 234 LLAGHDNTLA-----LGGLDLWOLF--QPDETPPGGALVFELW-----NRYVRVMY--OTM 280
DB 335 VTVGSDTHIALSSLLGVH-FHLPQYGADDPFGGALVGLWRTPDGQSVRAQYLAQSL 393

QY 281 DQLRNLPL-----LPRILPIPGCGSE--AACSLSDFA 312
DB 394 EQLRTVAPLDLDPHPLQQLTALGVCAQGMACPLPAFTQ 433

RESULT 15
ID Q8PF53 PRELIMINARY; PRT; 435 AA.
AC Q8PF53;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Acid phosphatase.
GN APPA OR XAC4132.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;

```

RN SEQUENCE FROM N.A.
RP STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Canarotte G., Cannavan F., Cardoso J., Chambergo F., Ciapina L.P.,
RA Cicaralli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Maidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.E., Tezza R.I.D.,
RA Spinoia L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities";
RL Nature 417:459-463(2002).
DR EXBL; AB012063; AM38967.1; -.
DR GO; GO:0003993; F:acid phosphatase activity; IEA.
DR InterPro; IPR000560; HisAc_phosphatase.
DR Pfam; PF00328; acid_phosphat; 1.
DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR Complete proteome.
SQ SEQUENCE 435 AA; 46186 MW; F85EC7B63F1511B7 CRC64;

Query Match 21.4%; Score 353; DB 16; Length 435;
Best Local Similarity 34.3%; Pred. No. 7e-19;
Matches 147; Conservative 39; Mismatches 116; Indels 126; Gaps 31;

Qy 4 LLAALLAALAPAAAA-AEPLKLVILSRHGVRSPTRKALMVSPD-----WPM- 52
Db 12 LLGLGLALVPLTPAQTADATPTDLEKVVILKGVHVR-----AMSSPERLCGRYSLHPHPF 67

Qy 53 --FVG--ITPRAALV--LIG--YR--YFARGLL--PGCPAAGTILADVDETR-RTGQAPAA 101
Db 68 GVPAGHLTANGAQLERLFGDYRYAHYATLGLNGDGCNQ-YYWANRTORTIASAQALAS 126

Qy 102 GLAPCAIVHAGDDPIFHGLDTGCLD-----QADAILAAGEGGLTARHG 145
Db 127 TLTFGCA-----DPVHH-VFTGSSDVLFDGTAALRQPQARMQAAIAGIGDAQAWN 179

Qy 146 LTLAKVLN-FASACLECAR-----VVGRLL-----GPLL-----ASTL 176
Db 180 ATQSDALDTLEHLLQLCAQPCPAQAPGKLRITVPAGLDDAGFSIPGLDGPAAASGI 239

Qy 177 SEIFLEVAQGP--MEVGMGRISAEWLL-----HNAQFLMN-RTPYIAR--ATPI---LI 223
Db 240 TESLMGWADGQDPFAALGNQGLDEATLURVPAPQAEFAURKRAPTVARLASTPLAARLL 299

Qy 224 VTAL-----SPARVLLAGHDNTLAL--GGDL-LWQLP--QPDETTPPGCALV 265
Db 300 ATLQCGNDHAHAEDAIGADARLVVVSGHGDTLTLLAGMFDLHWLPGVQSDQTVFGGALV 359

Qy 266 FELW-----NRYRVWY--QTMQDLRLNLEPLPRILP-----IPGCGSEAA-----CSLSD 309
Db 360 FERWRADGERVIRLRYTAQSLTQLRERVPLTLOAPPPSAIFIPGCGSTATPAYDCFLPQ 419

Qy 310 FARLVAPA 317
Db 420 FARIVQAA 427

```

Search completed: April 30, 2004, 12:41:28
Job time : 31.8084 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2004, 19:58:17 ; Search time 3538.44 Seconds
(without alignments)
16230.166 Million cell updates/sec

Title: US-10-021-723B-3

Perfect score: 1325

Sequence: 1 atgtcagattagaaatcg.....aacctgctgcacatctaa 1325

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sv.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.om.*

21: em.ov.*

22: em.pat.*

23: em.ph.*

24: em.pl.*

25: em.ro.*

26: em.sts.*

27: em.un.*

28: em.vi.*

29: em.htg.hum.*

30: em.htg.inv.*

31: em.htg.other.*

32: em.htg.mus.*

33: em.htg.pln.*

34: em.htg.rtd.*

35: em.htg.mam.*

36: em.htg.vrt.*

37: em.sv.*

38: em.htgo.hum.*

39: em.htgo.mus.*

40: em.htgo.pln.*

41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1325	100.0	1325	6	AX556791	AX556791 Sequence
2	1306	98.6	10029	1	AE013783	AE013783 Yersinia
3	1306	98.6	193050	1	AJ414149	AJ414149 Yersinia
4	1292.2	97.5	1325	6	AX556789	AX556789 Sequence
5	827.6	62.5	1326	6	AX556799	AX556799 Sequence
6	323	24.4	2558	1	AY378096	AY378096 Obesumbac
7	321.4	24.3	2161	1	AY425982	AY425982 Obesumbac
8	226.4	17.1	5421	6	AX042377	AX042377 Sequence
9	225.2	17.0	1901	1	ECOACPHD	L03373 Escherichia
10	225.2	17.0	1901	1	ECOACPHD	MS9708 Escherichia
11	225.2	17.0	1901	6	AX356570	AX356570 Sequence
12	225.2	17.0	1901	6	AX356571	AX356571 Sequence
13	225.2	17.0	1901	6	AX356572	AX356572 Sequence
14	224.6	17.0	3470	6	AX042376	AX042376 Sequence
15	224.6	17.0	4060	6	AX042374	AX042374 Sequence
16	224.6	17.0	6116	6	AX042375	AX042375 Sequence
17	224.6	17.0	6708	6	AX042373	AX042373 Sequence
18	224.6	17.0	10784	1	AE000200	AE000200 Escherich
19	224.6	17.0	17732	6	AX042378	AX042378 Sequence
20	224.6	17.0	18882	1	D90735	D90735 Escherichia
21	224.6	17.0	20623	6	AX042372	AX042372 Sequence
22	223.6	16.9	1901	1	ECOACPHA	L03370 Escherichia
23	223.6	16.9	1901	1	ECOACPHB	L03371 Escherichia
24	223.6	16.9	1901	6	AX356574	AX356574 Sequence
25	223	16.8	1315	1	AF537219	AF537219 Escherich
26	222	16.8	1901	1	ECOACPHC	L03372 Escherichia
27	222	16.8	1901	1	ECOACPHE	L03374 Escherichia
28	222	16.8	1901	1	ECOACPHF	L03375 Escherichia
29	218.8	16.5	1323	6	AR108133	AR108133 Sequence
30	218.8	16.5	1323	6	AR127818	AR127818 Sequence
31	218.8	16.5	1323	6	AR130956	AR130956 Sequence
32	218.8	16.5	1323	6	AR130956	AR130956 Sequence
33	218.8	16.5	1323	6	AX338966	AX338966 Sequence
34	218.8	16.5	1323	6	AX356566	AX356566 Sequence
35	218.2	16.5	1489	6	AR277753	AR277753 Sequence
36	218.2	16.5	10029	1	AE005292	AE005292 Escherich
37	218.2	16.5	32773	1	AP002554	AP002554 Escherich
38	215	16.2	11761	1	AE015127	AE015127 Shigella
39	215	16.2	292504	1	AE016981	AE016981 Shigella
40	211.8	16.0	301276	1	AE016758	AE016758 Escherich
41	177.4	13.4	1431	6	AX556797	AX556797 Sequence
42	143.6	10.8	1272	6	BD074747	BD074747 Novel phy
43	68.4	5.2	1675	1	ECOAGPA	M33807 E.coli peri
44	68.4	5.2	11458	1	AE015129	AE015129 Shigella
45	68.4	5.2	11710	1	AE000202	AE000202 Escherich

ALIGNMENTS

RESULT 1
AX556791
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AX556791
Sequence 3 from Patent WO0248332.
AX556791
AX556791.1 GI:258999903
Yersinia pestis
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Yersinia.

REFERENCE
AUTHORS
TITLE
JOURNAL

Short, J., Mathur, E.J., Richardson, T., Robertson, D. and Barton, N.
Recombinant phytases and uses thereof
Patent: WO 0248332-A 3 20-JUN-2002,

/note="residues 1 to 330 of 347 are 60.00 pct identical to residues 1 to 330 of 331 from GenPept : >gb|AA057489.1|AE005467.8 (AE005467) sucrose specific transcriptional regulator [Escherichia coli O157:H7 EDL933]"
 /codon_start=1
 /transl_table=11
 /product="putative sucrose specific transcriptional regulator"
 /protein_id="AA05371.1"
 /db_xref="GI:21958613"
 /translation="MVSLEKAVLASVLSMTASRAINSPEQLRPETYQVWQAIENLNVDPFAARKRGHATKSTLAVLAWDTATTFPSVETLLSELTAKEFGWNSFLVNLTS RPSORALNOLLACRDPGIIPTAGLRQVTIPEFRDKNIVLANCVSESHSIPTDPDQYQAKMLLQGGYQORPLCLVLPCTCKAGKARAGVEKAWQESALPTEQLNOCHIANGDHQVDPILPRLCRCPKGNFVILICGDDRIAFVLAQVLLAGVVIPOQVAVIG XGDMVGGEYLPLPTTQVLPYDYGQAALHIRGESAIRHLPCPLDERGSIKRS MPNNCVPLFSKQR"
 complement (1423..1683)
 /gene="y1804"
 CDS
 complement (1423..1683)
 /gene="y1804"
 /function="unknown"
 /note="residues 37 to 82 of 86 are 32.60 pct identical to residues 617 to 659 of 1018 from GenPept : >gb|AAK31959.1|AF285580.1 (AF285580) serine/threonine kinase 31 [Mus musculus]"
 /codon_start=1
 /transl_table=11
 /product="hypothetical"
 /protein_id="AA05372.1"
 /db_xref="GI:21958614"
 /translation="MSSINQTSVRAHTGETGALNPGEDHLIAHEALLVALMVECANES DNDQFLSIRETLNGILSSRYLDANAIALNGLIDEADKANR"
 complement (1803..3440)
 /gene="aer"
 /note="synonym: y1805"
 complement (1803..3440)
 /gene="aer"
 /function="regulator; degradation of small molecules; Carbon compounds"
 /note="residues 26 to 533 of 545 are 47.44 pct identical to residues 7 to 504 of 506 from E. coli K12 : B3072; residues 20 to 535 of 545 are 53.68 pct identical to residues 1 to 514 of 514 from GenPept : >emb|CAD18375.1| (AL646083) probable aerotaxis sensor receptor (chemotaxis transducer) transmembrane protein [Ralstonia solanacearum]"
 /codon_start=1
 /transl_table=11
 /product="aerotaxis sensor receptor, flavoprotein"
 /protein_id="AA05373.1"
 /db_xref="GI:21958615"
 /translation="MKYNGSIDYQGSPPFTTEVMNRVKNKPYRQYPIERTITLQSTTD IHGNIAYANAAPVRAAGFPEYQELQGPNNVHRPMPAFAFDMWQTNLAGLSSWILV KNRKNGDYVVRANATPLRNGRLTGIVRIAPTRDEVKQAEALYDFTNGAKRR HIALYRGILVETGMLSPSLFQTLPLRWLSALLSSAIIPTAASVMEVAGOPLLIL GSVALTVSLASLWLRQIARPIAIIQQAQDVSSGAGDYVQLNRVDEIGLMSVYN QLGNRLSTDVQGVGINTASNEIAGNRKLVKTEQATDNLQHVLSATGEQAAT VONSANANETTSLAEMSHAAKSGELMHQVETMGITNDSHRVIDIISVIEGIAP QTNILANAEVAAEQGQGFVAVASEVHLAQRSSTRAKEIKHLETYSVERVDRG STLVSQAGAMDNIVKQASQVSTLISEISTSTSTHQALQIQRSIRLDMQTYQNA MVEQYAGAAEBELHRTLRLTAAYRIYRLNNVGLRD"
 complement (4303..5160)
 /gene="y1806"
 CDS
 complement (4303..5160)
 /gene="y1806"
 /function="putative"
 /note="residues 1 to 279 of 285 are 65.59 pct identical to residues 1 to 279 of 286 from E. coli K12 : B4211; residues 1 to 282 of 285 are 65.60 pct identical to residues 1 to 282 of 284 from GenPept : >emb|CAB69475.1| (A92090) ORF1 [Pseudomonas sp.]"

/codon_start=1
 /transl_table=11
 /product="putative oxidoreductase"
 /protein_id="AA05374.1"
 /db_xref="GI:21958616"
 /translation="MIAVTGATGQLGRVIVNLVKKVPAHDIIAIVRSPDKAKDMLAL GYQVRADYNQPTTLDISAFKGVGKLLIISSEVQORVAQHTAVNAARAGVGLIAYT SILHADKSPGLGHEHRAETGLLRDSGLPVLLRNGWYTYENYASIAIPALAHGAFIGA AGRGFIASREDYAEAAATVVOENQAGKIYVLAGDDAYTLSEFSAITRQSGKAVA YQDLSEADFKHALISAGLPDGLAGLADSDAGAAQGLFDDSHLTKLIGRTTTPYAK VITATLATR"
 5131..5703
 /gene="y1807"
 CDS
 5131..5703
 /gene="y1807"
 /note="residues 68 to 189 of 190 are 62.29 pct identical to residues 34 to 155 of 156 from E. coli K12 : B4212"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AA05375.1"
 /db_xref="GI:21958617"
 /translation="MAGTGDNDNDAPVLEEOLTLREDFGYICPKSSLPKSVAYI SLIVSTYKXSVRVYTKLQNMGNISFPQELRGELLVNDPCSRTELKVTSGWGLVL IALSNETLRFSAIRKIGGVSEKMLAQTLQHLBEDGFDIDRTAYFVVPFVHSYSLTPLG KBIQOVLEGLTVWLETNLHRIMEORQQRSV"
 5785..5934
 /gene="y1808"
 CDS
 5785..5934
 /gene="y1808"
 /function="unknown"
 /note="residues 6 to 49 of 49 are 42.22 pct identical to residues 61 to 101 of 380 from GenPept : >dbj|BAE70104.1| (AP002924) cytochrome b [Lampiris guttatus]"
 /codon_start=1
 /transl_table=11
 /product="hypothetical"
 /protein_id="AA05376.1"
 /db_xref="GI:21958618"
 /translation="MLEGTAPDVTLPVPTIGQARFGLRHWLKSTAPTAGPFFICIYL SAARG"
 complement (5931..6023)
 /gene="y1809"
 CDS
 complement (5931..6023)
 /gene="y1809"
 /function="unknown"
 /note="residues 1 to 26 of 30 are 34.61 pct identical to residues 16 to 41 of 444 from GenPept : >gb|AAK78422.1|AE007558.12 (AE007558) Permease, putative chloride channel [Clostridium acetobutylicum]"
 /codon_start=1
 /transl_table=11
 /product="hypothetical"
 /protein_id="AA05377.1"
 /db_xref="GI:21958619"
 /translation="MSFTWVILGVVLCVIGALLKIFGKKKDR"
 6235..7560
 /gene="appa"
 /note="synonym: y1810"
 /gene="appa"
 6235..7560
 /function="enzyme; central intermediary metabolism: Pool, multipurpose conversions"
 /note="pH 2.5 acid phosphatase; periplasmic; residues 13 to 441 of 441 are 45.87 pct identical to residues 5 to 432 of 432 from E. coli K12 : B0980"
 /codon_start=1
 /transl_table=11
 /product="phosphoanhydride phosphorylase"
 /protein_id="AA05378.1"
 /db_xref="GI:21958620"
 /translation="MSVLENRVRLSGLVLMGLAAITAPVAAPSGYTLERVVILSR HGVRSTKQTQLMNDVTPDKWQWPVKAQVLTLPARGAELVTLMGGFYDFRSLGLIAA

GCPAEGGVYAQADIDQRLTGTQAFDVGAPGGLTVHNDLKEKTDLFHPVEAGVC
 KLDAAQTDRAIEBQLGPDLDTSQRYAKPFAQMGVDFNFAASPYCKSLQOQKTCDFP
 HFAENVNVEKGTQVTLSPGLALSITLGEIFLQNAQAPVAVWCRLLKRGWYVSL
 SLHNAQFNMAKTPYTLARHGKPLLOQIDTALTLODAGOKLP:SAONRVLFGLGHD
 TNANTAGMLGDMOLPEOPDNTPPGGGLVFLMQLNPNHORYVAVKPYQTMQDLRN
 AEKLDKNPAGLISVAVACENNGDKLCELDTPKRVKRVIEPACHI"
 complement (7862..8077)
 /gene="y1811"
 CDS
 complement (7862..8077)
 /gene="y1811"
 /notes="residues 9 to 71 of 71 are 41.26 pct identical to
 residues 4 to 66 of 67 from E. coli K12 : B1541; residues
 6 to 70 of 71 are 44.61 pct identical to residues 1 to 65
 of 67 from Genpept : >emb|CAD09963.1| (AUS13384)
 hypothetical protein [Salmonella enterica subsp. enterica]

Query Match 98.6%; Score 1306; DB 1; Length 10029;
 Best Local Similarity 99.5%; Pred. No. 0;
 Matches 1320; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY	1	ATGTCAGTATTAGAAAATCGTGTACGGCTATCTGGCTGGTATTGATGTAAGCGGATTG	60
DB	6235	ATGTCAGTATTAGAAAATCGTGTACGGCTATCTGGCTGGTATTGATGTAAGCGGATTG	6294
QY	61	GCTGCTATTACTGCGCGGTAGCGCGGAGCCATCGGGCTATCTTTAGAACTGTGGTT	120
DB	6295	GCTGCTATTACTGCGCGGTAGCGCGGAGCCATCGGGCTATCTTTAGAACTGTGGTT	6354
QY	121	ATTTGAGTCGCATGCTGCTGCTCGCGACCAAC-AACGAGCTTATGATGATGTT	179
DB	6355	ATTTGAGTCGCATGCTGCTGCTCGCGACCAACAAACGAGCTTATGATGATGTT	6414
QY	180	ACGCCAGATAAGTGGCCACAAATCGCGGTAAAGCGGGTATTTCAGCCGACGCTGTGCG	239
DB	6415	ACGCCAGATAAGTGGCCACAAATCGCGGTAAAGCGGGTATTTCAGCCGACGCTGTGCG	6474
QY	240	GAGTGTGTCATGATGCGGGGGTTTTATGTTGATTAATTTGCGAGCTTGGTTGTTA	299
DB	6475	GAGTGTGTCATGATGCGGGGGTTTTATGTTGATTAATTTGCGAGCTTGGTTGTTA	6534
QY	300	GCGCGGGGATGTCGCGCAGAGGGGGGTATATGCACGACGATATCGATCAACGTACC	359
DB	6535	GCGCGGGGATGTCGCGCAGAGGGGGGTATATGCACGACGATATCGATCAACGTACC	6594
QY	360	CGTTTAAACCGGACAGGCATTTCTTGATGTGTGCTCGGGGTGTTGGTTGACCGTGCAT	419
DB	6595	CGTTTAAACCGGACAGGCATTTCTTGATGTGTGCTCGGGGTGTTGGTTGACCGTGCAT	6654
QY	420	AATCAGCGCATTTGAAAGACCGATCCCTGTTCCATCCGTAGACGACTGCGGTGT	479
DB	6655	AATCAGCGCATTTGAAAGACCGATCCCTGTTCCATCCGTAGACGACTGCGGTGT	6714
QY	480	AAGTTAGACAAACACAAACAGATAAAGCGATTGAAGACGATTGGCGGGCGGTTAGAT	539
DB	6715	AAGTTAGACCGCCACAAACAGATAAAGCGATTGAAGACGATTGGCGGGCGGTTAGAT	6774
QY	540	ACGGTGAAGCCAGCTACGTAAACCTTTTCCAGATGGGACGCTGCTGAATTTGCG	599
DB	6775	ACGGTGAAGCCAGCTACGTAAACCTTTTCCAGATGGGACGCTGCTGAATTTGCG	6834
QY	600	GCTTCTCTTATTGCAAACTTTGCAACAGCAAGGAAACGTTGATTTTGGCCACTTT	659
DB	6835	GCTTCTCTTATTGCAAACTTTGCAACAGCAAGGAAACGTTGATTTTGGCCACTTT	6894
QY	660	GCGGCCAATGAAGTTACGTTATTAAGAGGACAAAGTGAACCTCAGTGGGCCACTG	719
DB	6895	GCGGCCAATGAAGTTACGTTATTAAGAGGACAAAGTGAACCTCAGTGGGCCACTG	6954
QY	720	GCGTTATCATCGACATTTGGGTGAATCTTTATTACAAACCGCAAGCCATGCCAGAG	779
DB	6955	GCGTTATCATCGACATTTGGGTGAATCTTTATTACAAACCGCAAGCCATGCCAGAG	7014
QY	780	GTTGCTCGGCAACGGCTAAAGGGGGGAGAAATGGGTATTCCTTATTGTCATTACATAAC	839

DB	7015	GTTCCTGGCAACCGCTAAAAAGGGCGGAGAAATGGGTATCTTATTGTCTATTACATAAC	7074
QY	840	GCGCAATTTAATTAATGCGAAACACCGCTATATCCCGTCTATAAAGGAGCGCCATT	899
DB	7075	GCGCAATTTAATTTAATGCGAAACACCGCTATATCCCGTCTATAAAGGAGCGCCATT	7134
QY	900	TTACAGCAGATAGATACGGCTTTTAAACCTTTCAACTGGATGCTCAGGGGCAAAAGCTACCC	959
DB	7135	TTACAGCAGATAGATACGGCTTTTAAACCTTTCAACTGGATGCTCAGGGGCAAAAGCTACCC	7194
QY	960	ATTTACGCCCAAAACCGGGCTTTGTTCTCGTGGGATGATACCAATATTTGCCAATATT	1019
DB	7195	ATTTACGCCCAAAACCGGGCTTTGTTCTCGTGGGATGATACCAATATTTGCCAATATT	7254
QY	1020	GCGGGTATGTCGGAGCGGACTGCGACCTACCCGAGCAACCTGATATATCTCCACCAAGT	1079
DB	7255	GCGGGTATGTCGGAGCGGACTGCGACCTACCCGAGCAACCTGATATATCTCCACCAAGT	7314
QY	1080	GCGGGATTTGTTTGAATATGCGAAATCCCGATACCAACGCGCTACGTTGCGGTG	1139
DB	7315	GCGGGATTTGTTTGAATATGCGAAATCCCGATACCAACGCGCTACGTTGCGGTG	7374
QY	1140	AAGATGTTTACCAACGATGGATCAGTTGCGTAAATCCGAAATAATTTGGATCTGAAATAAT	1199
DB	7375	AAGATGTTTACCAACGATGGATCAGTTGCGTAAATCCGAAATAATTTGGATCTGAAATAAT	7434
QY	1200	AACCCAGGGGTATTTTCCGTTGCGTGTGCTGGTGTGTAATAACGGTACGATAG	1259
DB	7435	AACCCAGGGGTATTTTCCGTTGCGTGTGCTGGTGTGTAATAACGGTACGATAG	7494
QY	1260	CTTTGCGAGCTTGATATCAATCCAAAGAAAGTGGCTAAAGTAAATTTGAACCTGCTGCCAC	1319
DB	7495	CTTTGCGAGCTTGATATCAATCCAAAGAAAGTGGCTAAAGTAAATTTGAACCTGCTGCCAC	7554
QY	1320	ATCTAA 1325	
DB	7555	ATCTAA 7560	

RESULT 3
 AJ414149
 LOCUS
 DEFINITION
 Yersinia pestis strain CO92 complete genome; segment 9/20.
 ACCESSION
 AJ414149.1 GI:15979570
 VERSION
 AJ414149.1
 KEYWORDS
 Yersinia pestis CO92
 SOURCE
 Yersinia pestis CO92
 ORGANISM
 Yersinia pestis CO92
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Yersinia.
 REFERENCE
 1 (bases 1 to 193050)
 Parkhill, J., Wren, B.W., Thomson, N.R., Titball, R.W., Holden, M.T.G.,
 Prentice, M.B., Sebaihia, M., James, K.D., Churcher, C., Mungall, K.L.,
 Baker, S., Basham, D., Bentley, S.D., Brooks, K., Cerdano-Tarraga, A.M.,
 Chillingworth, T., Cronin, A., Davies, R.M., Davis, P., Dougan, G.,
 Feltwell, T., Hamlin, N., Holroyd, S., Jagers, K., Leather, S.,
 Karyshev, A.V., Moule, S., Oyston, P.C.F., Quail, M., Rutherford, K.,
 Simmonds, M., Skelton, J., Stevens, K., Whitehead, S. and Barrall, B.G.
 Genome sequence of Yersinia pestis, the causative agent of plague
 Nature 413 (6855), 523-527 (2001)
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 21470413
 11586360
 2 (bases 1 to 193050)
 Parkhill, J.
 Direct Submission
 Submitted (04-OCT-2001) Submitted on behalf of the Yersinia
 sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
 Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
 JOURNAL
 COMMENT
 Notes:
 Details of Y. pestis sequencing at the Sanger Centre are available
 on the World Wide Web.
 (URL, http://www.sanger.ac.uk/Projects/Y_pestis/).

FEATURES	source	misc_feature	gene	CDS	misc_feature	gene	CDS
Location/Qualifiers							
1. 193050							
/organism="Versinia pestis CO92"							
/mol_type="genomic DNA"							
/strain="CO92"							
/db_xref="taxon:214092"							
/note="biovar: Orientalis"							
372. .1151							
/gene="YPO1528"							
372. .1151							
/gene="YPO1528"							
/note="Similar to Escherichia coli ferric iron reductase protein FhuF SW:PHUF_ECOLI (P39405) (262 aa) fasta scores: E(): 8.4e-33, 35.0% id in 246 aa"							
/codon_start=1							
/transl_table=11							
/product="putative ferric iron reductase"							
/protein_id="CAC90351.1"							
/db_xref="GI:15979571"							
/db_xref="GOA:Q8ZF25"							
/db_xref="SPTREMBL:Q8ZF25"							
/translation="MSNTAQITLTPSGIADIATPELPEKTFAPHSRTLKNVADIDPEE TMSFHTSSINFTPLTKYDRDEYDGNLDKPNDKALYSWSQWYGLIIPPMWLLLI EYPOTIDHKNFVLFHSGRPVYVYOLKWSQDPGTLLELYILLNHVPIPAEK IESVQGLNGRLWNIGVLMFWYLGEEKRGLGDDLYQSLINGLFMELSLPQGDNPLY RTVMRNGTQLRRSCCRNKLPGVRSCHDCPLEKPPNLLIS"							
1303. .2850							
/gene="YPO1529"							
1303. .2850							
/gene="YPO1529"							
/note="Similar to Enterobacter aerogenes L-2,4-diaminobutyrate decarboxylase ddc TR:Q980P8 (EMBL:AB02468) (490 aa) fasta scores: E(): 0, 39.3% id in 468 aa, and to Acinetobacter baumannii L-2,4-diaminobutyrate decarboxylase ddc SW:DDC_ACIBA (Q43908) (510 aa) fasta scores: E(): 0, 38.3% id in 431 aa"							
/codon_start=1							
/transl_table=11							
/product="putative decarboxylase"							
/protein_id="CAC90352.1"							
/db_xref="GI:15979572"							
/db_xref="GOA:Q8ZF24"							
/db_xref="SPTREMBL:Q8ZF24"							
/translation="MYKDIYEENHFIDQGNIDTDVLCLENFLANKDSAEHYEVNWNIT ARIKAVTLVDKPSGVLPSLELSPDFCIDLNQPLSILDSALDELQQLYIKDAIYFHH PKYVAHNCVPVLPVLAELSIITSNLSLDTPQSGATLIEQKVLDDTRDKWALGSGQ GDGIFTSGGTQSNMALLARDHPCORDANQKQLGQLPADSHKFRIFTSVVSHFST QKAAALGLGYNVAVPYDSEFMDQALDQSLCSQGNIPLAIVTGTGTFDGS IDPLHTISTAKOYGLWLYDAAYCGGLVSHYEQLLSGIEMADSVTVYHKSFFQP VSCAPFVKQHPSTLYTHAEYLNPLSLASLEPNLVNKSIOITFRFDALKLWLLP IMGAKQLGAPRHVMDTAQAYQLMLNPYFELTHRPVLTSLVFRWPEQPSDDID RTNAYIRKALPREGHTVIAGTKYNGKQYLKFTPLNPTTLTQLHNEIVDALITGKNYP TLVNSQS"							
1699. .2619							
/gene="YPO1529"							
/note="Pfam match to entry PF00282 pyridoxal deC, Pyridoxal-dependent decarboxylase conserved domain, score 165.20, E-value 2.7e-47"							
2885. 4183							
/gene="alca"							
/note="synonym: YPO1530"							
2885. 4183							
/gene="alca"							
/note="Similar to Bordetella bronchiseptica alcaligin biosynthesis enzyme SW:ALCA_BORBR (Q44740) (461 aa) fasta scores: E(): 0, 49.6% id in 425 aa, and to Streptomyces coelicolor putative monooxygenase Scc105.14 TR:Q9L071 (EMBL:AL163641) (413 aa) fasta scores: E(): 0, 53.7% id in 419 aa"							
/codon_start=1							
/transl_table=11							
/product="putative siderophore biosynthetic enzyme"							
protein_id="CAC90353.1"							
/db_xref="GI:15979573"							
/translation="MRPYDFIGIGPFLNLACLTQPLNDVCCFLIDQKAGFDWHPGC MMLESATLQTPFMSDLVTLASPELSELYIKQQRIYSFYRENFLLRKEYNQYC QWAVEQLNLRNTQVDIHYDEDTALRYLTCSDTCSGRHEFLCKLCEIDRFDPSPHI PECCOFFIOQIVMSGOYLEHVALQKRAITLIGSGAAEFYLLCLCEIDRFDPQLN WLTRSPRELPLEYTKLTEMSPDYVYFHLPEPKRDELNLTHKPLVKGINGTILND IFDLYTKELGALANVLYTNSSELKYSVSHDGLHLVSLVQHEQNLQFTEGAVL ATGIHYQPTTFIEGRQHRIMDNKGRIVQVNSIDRNNTIFVQNVELHFGVITPDL GMACYRNSCLIREITGTBYPIEKNIAQFQSAPSLRRKE"							
2912. .2944							
/gene="alca"							
/note="PS00013 Prokaryotic membrane lipoprotein lipid attachment site."							
4180. 4761							
/gene="alcb"							
/note="synonym: YPO1531"							
4180. 4761							
/gene="alcb"							
/note="Similar to Bordetella bronchiseptica Alcb alcb TR:P94254 (EMBL:U61153) (201 aa) fasta scores: E(): 2e-27, 43.7% id in 174 aa, and to Streptomyces coelicolor putative acetyltransferase scc105.15 TR:Q9L070 (EMBL:AL163641) (184 aa) fasta scores: E(): 3e-23, 39.5% id in 172 aa"							
/codon_start=1							
/transl_table=11							
/product="putative siderophore biosynthetic enzyme"							
/protein_id="CAC90354.1"							
/db_xref="GI:15979574"							
/db_xref="SPTREMBL:Q8ZF22"							
/translation="MMNARNIESDVTVRSGFQRLRPLVLAHDVTLMDWNVREYAHY WMGQSQVQKAFYQELTQPGGVYVYQQQPAFLLEYLAQHDVIADYDPAQPD DVMGHILVAPVERITGTWRVQAIPTFISDPTVARIVVEPDVNEKIHRLNKRAG FVYQQLTLPHKTAWLAFCTREQVRLAQKEHQ"							
4758. 6602							
/gene="YPO1532"							
4758. 6602							
/gene="YPO1532"							
/note="Similar to Bordetella bronchiseptica Alcc TR:P94255 (EMBL:U61153) (618 aa) fasta scores: E(): 0, 53.5% id in 608 aa, and to Streptomyces coelicolor hypothetical 66.6 kDa protein Scc105.16 TR:Q9L069 (EMBL:AL163641) (595 aa) fasta scores: E(): 0, 49.4% id in 589 aa. Similar to YPO1538 (46.8% identity in 596 aa overlap)"							
/codon_start=1							
/transl_table=11							
/product="putative siderophore biosynthetic enzyme"							
/protein_id="CAC90355.1"							
/db_xref="GI:15979575"							
/db_xref="GOA:Q8ZF21"							
/db_xref="SPTREMBL:Q8ZF21"							
/translation="MMDLAPLSHQSLPHLQPLVMQKNRHLHLCALSBFSHEC LLAPQMTDHPDSGYDYQLVAAADKPNYVFRARLADHLIDLDPDSLHKTVNEK STDLDLLLFIIEFKQLSISERVLPYLTLEETSTLYSAFKHCRITGIGATLVNASFQ IIEKMEGHPFSFVANNRIGFDAQDFORFSPAASDVHLVLAHAKSKAHFACIAEQ DYAKMEGELGAEVLARFEQOLIARDQNPQYVHPQWQNKLTISIADIANQR LVPLQGDQAYQAQOSIRTFNRSHPQRYVYKMAISLLMGFMGLSPYKWTTPGIN EWLFDVGEDEILOAYDFKILREVASIGFRNSYEQATIGTSAYKMAAALWRENPLQ LIQPCQLMTWALHLDQEGKALLPALIDSLGUSTQQLWRLYSKSLPLHLCLDYV DLMFMFGENTILVQNAVPOHIFMKDLAEILVLTNEADLPFKARRIIVMDPDEMKT LTILSDVDFGVFVLAALVDQGDYPERYFVADVAACIRQYQQQYFELAAKFAF5YPLF TPXIKRCCNLRLNNQHMLDLADPVQSQFADDLINPLVEFEVSAG"							
complement(6711..7559)							
/gene="YPO1533"							
complement(6711..7559)							
/gene="YPO1533"							
/note="Similar to Vibrio cholerae ferric vibriobactin ABC transporter, ATP-binding protein vc0779 TR:Q9KTV8 (EMBL:AB004163) (279 aa) fasta scores: E(): 0, 55.9% id in 263 aa, and to Streptomyces coelicolor putative iron-siderophore uptake system ATP-binding component							

```
sci51.25C TR:Q9S215 (EMBL:AL109848) (301 aa) fasta scores:
E(): 0, 57.2% id in 264 aa"
/codon_start=1
/transl_table=1
/product="putative iron-siderophore transport system,
ATP-binding component"
/protein_id="CAC90356.1"
/db_xref="GI:15979576"
/db_xref="GOA:Q8ZP20"
/db_xref="SPTREMBL:Q8ZP20"
/translation="WRSQRELTSPNPLENGEDISDKPRLTQKLSLGYGOOLINKL
SUSIKNGAFSVIIGNCGSKLLRALSRLLPQNGSRILQDQDIOHYKAKVFAQUS
LUSQASSETUTVDFLSVRGHAHCFHFFHSTEDERIVRAALSSQVSLVOORVS
ELSGQRQAFAMTIAQHTPIVLDDEPTIYDIAHETIMDLCOELQOGLKVLVL
HDINQALRYATHLIMLKQIYAELGPESIVTSDSIAAFVGLRCHRIIDPDSGRPLVI
PRQQR"
complement(6849..7409)
/gene="YPO1533"
/notes="Fram match to entry PF00005 ABC_tran, ABC
transporter, score 183.50, E-value 3.5e-51"
complement(7032..7076)
/gene="YPO1533"
/notes="PS00211 ABC transporters family signature."

Query Match          98.6%; Score 1306; DB 1; Length 193050;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1320; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 1 ATGTCAGTATTAGAAATCGTGACGGCTATCTGGCTGGTATTGATGCTAAGCGATTG 60
Db 135753 ATGTCAGTATTAGAAATCGTGACGGCTATCTGGCTGGTATTGATGCTAAGCGATTG 135812

QY 61 GTCGCTATTACTGCGCGGTAGCGCGGAGCCATCGGCTATACCTTTAGAACTGTGTT 120
Db 135813 GTCGCTATTACTGCGCGGTAGCGCGGAGCCATCGGCTATACCTTTAGAACTGTGTT 135872

QY 121 ATTTTGAGTCGCCATGGTGTTCCTCGCGGACCAAAAC-AACGAGCTTATGATGATGTT 179
Db 135873 ATTTTGAGTCGCCATGGTGTTCCTCGCGGACCAAAACGAGCTTATGATGATGTT 135932

QY 180 ACCGACAGTAAGTGGCCACATGCGCGGTAAAGCGGGTATTAAAGCCACGTTGGTGG 239
Db 135933 ACCGACAGTAAGTGGCCACATGCGCGGTAAAGCGGGTATTAAAGCCACGTTGGTGG 135992

QY 240 GAGTTGTCACATTGATGGGGGGTATTATGGTATTACTTTGCGAGCCTGGTTGTGTTA 299
Db 135993 GAGTTGTCACATTGATGGGGGGTATTATGGTATTACTTTGCGAGCCTGGTTGTGTTA 136052

QY 300 GCGCGGGGATGTCGCGCAGAGGGGGGGTATATGCAAGCAGATATCGATCAACGTACC 359
Db 136053 GCGCGGGGATGTCGCGCAGAGGGGGGGTATATGCAAGCAGATATCGATCAACGTACC 136112

QY 360 CGCTTAACCGACAGGCAATTTCTTGATGGTGTGGCTCCGGGGTGTGGTTGACCGTGCAT 419
Db 136113 CGCTTAACCGACAGGCAATTTCTTGATGGTGTGGCTCCGGGGTGTGGTTGACCGTGCAT 136172

QY 420 AATCAGCCGATTTGAAAAAGACCGATCCCTGTTCCATCCGCTAGAGACTGGCGTGTGT 479
Db 136173 AATCAGCCGATTTGAAAAAGACCGATCCCTGTTCCATCCGCTAGAGACTGGCGTGTGT 136232

QY 480 AGTTTAGACACGACACACAGATAGACGATTAAGACGATTGGCGGGCGGTAGAT 539
Db 136233 AGTTTAGACGCGCACACACAGATAGACGATTAAGACGATTGGCGGGCGGTAGAT 136292

QY 540 ACGGTGACCGACGCTACGTAAACCTTTTGCCAGATGGGACGCTGTAATTTGCG 599
Db 136293 ACGGTGACCGACGCTACGTAAACCTTTTGCCAGATGGGACGCTGTAATTTGCG 136352

QY 600 GCTTCTCCTTATTGCAATCTTTGACACAGAGGAAACGTTGATTTTGGCCACTTT 659
Db 136353 GCTTCTCCTTATTGCAATCTTTGACACAGAGGAAACGTTGATTTTGGCCACTTT 136412

QY 660 GCGGCCAATGAAGTAACTGTTAATAAAGAGGACAAAAGTACCCCTCAGTGGGCCACTG 719
```

```
Db 136413 GCGGCCAATGAAGTAACTGTTAATAAAGAGGACAAAAGTGACCTCAGTGGGCCACTG 136472
QY 720 GCGTTATCATCGCATTTGGTGAATCTCTTATTACAAAACGCAAGCCATGCCAGAG 779
Db 136473 GCGTTATCATCGCATTTGGTGAATCTCTTATTACAAAACGCAAGCCATGCCAGAG 136532
QY 780 GTTGCCCTGGCACCGGCTAAAAGGGGGGGGAATTTGGTATCCTTATTGTCTATTACATAAC 839
Db 136533 GTTGCCCTGGCACCGGCTAAAAGGGGGGGGAATTTGGTATCCTTATTGTCTATTACATAAC 136592
QY 840 GCGCAATTTAATTAATGCAAAAACACCGTATATCGCGCTCATAAAGGAGCGCCATTA 899
Db 136593 GCGCAATTTAATTAATGCAAAAACACCGTATATCGCGCTCATAAAGGAGCGCCATTA 136652
QY 900 TTACAGCAGATAGATACGGCTTTTAAACCTTTCAACCTGATGCTCAGGGGCAAAAGCTACCC 959
Db 136653 TTACAGCAGATAGATACGGCTTTTAAACCTTTCAACCTGATGCTCAGGGGCAAAAGCTACCC 136712
QY 960 ATTTACGCCCAAAACCGGCTTTTGTCTCGGTGGGCGATGATACCAATATTGCCAATATT 1019
Db 136713 ATTTACGCCCAAAACCGGCTTTTGTCTCGGTGGGCGATGATACCAATATTGCCAATATT 136772
QY 1020 GCGGCTATGCTGGGAGCGGACTGCGCAGCTACCCGAGCAACCTGATATAATCTCCACGAGGT 1079
Db 136773 GCGGCTATGCTGGGAGCGGACTGCGCAGCTACCCGAGCAACCTGATATAATCTCCACGAGGT 136832
QY 1080 GGGGGATTGGTTTTTGAACCTATGCGCAAAATCCCGATTAACACGAGCGCTACGTTGGGTG 1139
Db 136833 GGGGGATTGGTTTTTGAACCTATGCGCAAAATCCCGATTAACACGAGCGCTACGTTGGGTG 136892
QY 1140 AAGATGTTTACCAACGATGGATCAGTTGCGTAAATGCCGAAAAATTTGGAATCTGAAAAAT 1199
Db 136893 AAGATGTTTACCAACGATGGATCAGTTGCGTAAATGCCGAAAAATTTGGAATCTGAAAAAT 136952
QY 1200 AACCCAGCGGATATTATTTCCGTTGCGTTGCGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1259
Db 136953 AACCCAGCGGATATTATTTCCGTTGCGTTGCGTTGTTGTTGTTGTTGTTGTTGTTGTTG 137012
QY 1260 CTTTTCGAGCTTGATACATTCACAAAGAAAGTGGCTAAAGTAAATTTGAACCTGCTGCCAC 1319
Db 137013 CTTTTCGAGCTTGATACATTCACAAAGAAAGTGGCTAAAGTAAATTTGAACCTGCTGCCAC 137072
QY 1320 ATCTAA 1325
Db 137073 ATCTAA 137078

RESULT 4
AX556789
LOCUS
DEFINITION Sequence 1 from Patent WO0248332.
ACCESSION AX556789
VERSION AX556789.1 GI:25899902
KEYWORDS Yersinia pestis
SOURCE Yersinia pestis
ORGANISM Yersinia pestis
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Yersinia.
REFERENCE
1 Short J., Mathur, E.J., Richardson, T., Robertson, D. and Barton, N.
Recombinant phages and uses thereof
Patent: WO 0248332-A 1 20-JUN-2002;
DIVERSA CORPORATION (US)
FEATURES
source
1..1325
/organism="Yersinia pestis"
/mol_type="unassigned DNA"
/db_xref="taxon:632"

ORIGIN
Query Match 97.5%; Score 1292.2; DB 6; Length 1325;
Best Local Similarity 99.2%; Pred. No. 0;
```

Matches 1316; Conservative 0; Mismatches 8; Indels 2; Gaps 2;
 QY 1 ATGTCAGTATTAGAAATCGGTACGGCTATCTCGGTGTTATGATGCTTAAGCGGATG 60
 Db 1 ATGTCAGTATTAGAAATCGGTACGGCTATCTCGGTGTTATGATGCTTAAGCGGATG 60
 QY 61 GCTGCTATTACTGCGCGGTAGCGCGGAGCCATCGGCTATACCTTAGAACGTTGGTT 120
 Db 61 GCTGCTATTACTGCGCGGTAGCGCGGAGCCATCGGCTATACCTTAGAACGTTGGTT 120
 QY 121 ATTTTGGTCCGATGTTGCTCGCCGACCAAC-ACGCGAGCTTATGATGATGTT 179
 Db 121 ATTTTGGTCCGATGTTGCTCGCCGACCAAC-ACGCGAGCTTATGATGATGTT 180
 QY 180 ACGCCAGATAAGTGGCCACAAATGCGCGGTAAAGCGGGTATTTAAACGCGAGTGGG 239
 Db 181 ACGCCAGATAAGTGGCCACAAATGCGCGGTAAAGCGGGTATTTAAACGCGAGTGGG 240
 QY 240 GAGTTGGTCAATGATGGGGGGTATATGGTGAATTAATTCGACGCTTGGTTGTTA 299
 Db 241 GAGTTGGTCAATGATGGGGGGTATATGGTGAATTAATTCGACGCTTGGTTGTTA 300
 QY 300 GCGCGGGGATGTCGCGCAGAGGGGGGTATATGACAGGAGATATCGATCAACGTACC 359
 Db 301 GCGCGGGGATGTCGCGCAGAGGGGGGTATATGACAGGAGATATCGATCAACGTACC 360
 QY 360 CGTTTACCGGACAGGCAATTTCTTGATGGTGGCTCGGGGTGTGGTTGACCGTGCAT 419
 Db 361 CGTTTACCGGACAGGCAATTTCTTGATGGTGGCTCGGGGTGTGGTTGACCGTGCAT 420
 QY 420 AATCAGGCGGATTTGAAAGACCGATCCCTGTTTCCATCCGTAGAGACTGGGCTGT 479
 Db 421 AATCAGGCGGATTTGAAAGACCGATCCCTGTTTCCATCCGTAGAGACTGGGCTGT 480
 QY 480 AAGTTAGACACGACCAAAAGATGAGAAAGATGAGAAAGATGAGAAAGATGAGAAAG 539
 Db 481 AAGTTANACGNNGCACAAACAGATAAGAGATGAGAAAGATGAGAAAGATGAGAAAG 540
 QY 540 ACGGTGACCGAGCTACGCTAAACCTTTCCAGAGTGGGGAGCTGCTGAATTTGCG 599
 Db 541 ACGGTGACCGAGCTACGCTAAACCTTTCCAGAGTGGGGAGCTGCTGAATTTGCG 600
 QY 600 GCTTCTCTTATGCAATCTTTGCAACAGCAAGAAAGATGAGAAAGATGAGAAAGATG 659
 Db 601 GCTTCTCTTATGCAATCTTTGCAACAGCAAGAAAGATGAGAAAGATGAGAAAGATG 659
 QY 660 GCGGCAATGAGTTAAAGTTAAAGAGGAGAAAGTGGCTGAGTGGGCGGCACTG 719
 Db 660 GCGGCAATGAGTTAAAGTTAAAGAGGAGAAAGTGGCTGAGTGGGCGGCACTG 719
 QY 720 GGTATCATGACATGCGGTGAATCTTTTATTAACAAAGCAAGCAAGCAAGCAAGCAAG 779
 Db 720 GGTATCATGACATGCGGTGAATCTTTTATTAACAAAGCAAGCAAGCAAGCAAGCAAG 779
 QY 780 GTTGCTGCAAGGCTAAAGGGGCGGAGATTTGGGTATCTTTATTTGCTATACATAAC 839
 Db 780 GTTGCTGCAAGGCTAAAGGGGCGGAGATTTGGGTATCTTTATTTGCTATACATAAC 839
 QY 840 GCGCAATTTAATTTGCAAAACACCGTATATGCGCGGTATATAAGGAGCGCAATTA 899
 Db 840 GCGCAATTTAATTTGCAAAACACCGTATATGCGCGGTATATAAGGAGCGCAATTA 899
 QY 900 TTACAGCAGATAGATACGCTTTAAACCTTTGATGCTGATGCTGAGGCGCAAGTACCC 959
 Db 900 TTACAGCAGATAGATACGCTTTAAACCTTTGATGCTGATGCTGAGGCGCAAGTACCC 959
 QY 960 ATTTGAGCCAAACCGGGTCTTTGCTCGGTGGGCAATGATACCAATTTGCCAATATT 1019
 Db 960 ATTTGAGCCAAACCGGGTCTTTGCTCGGTGGGCAATGATACCAATTTGCCAATATT 1019
 QY 1020 GCGGGTATGCTGGGAGCGAGTGGAGCTTACCGGAGCAACCTGATTAATCTCCACAGGT 1079
 Db 1020 GCGGGTATGCTGGGAGCGAGTGGAGCTTACCGGAGCAACCTGATTAATCTCCACAGGT 1079

QY 1080 GGGGATTTGGTTTGAACATATGGCAAAATCCGATAACCAACGCGGTACGTTGCGGTG 1139
 Db 1080 GGGGATTTGGTTTGAACATATGGCAAAATCCGATAACCAACGCGGTACGTTGCGGTG 1139
 QY 1140 AAGATGTTTACCAACGATGATCAGTTGCGTAAATGCGAAATTTGATGATGAAAT 1199
 Db 1140 AAGATGTTTACCAACGATGATCAGTTGCGTAAATGCGAAATTTGATGATGAAAT 1199
 QY 1200 AACCAGCGGGTATTATTTCCGTTGCGTAAATGCGTAAATGCGTAAATGCGTAAAT 1259
 Db 1200 AACCAGCGGGTATTATTTCCGTTGCGTAAATGCGTAAATGCGTAAATGCGTAAAT 1259
 QY 1260 CTTTGGAGCTTGTATACATTCATAAAGAAAGTGGCTAAATTTGAACCTGCTGCCAC 1319
 Db 1260 CTTTGGAGCTTGTATACATTCATAAAGAAAGTGGCTAAATTTGAACCTGCTGCCAC 1319
 QY 1320 ATCTAA 1325
 Db 1320 ATCTAA 1325

RESULT 5
 AX556799
 LOCUS 1326 bp DNA linear PAT 27-NOV-2002
 DEFINITION Sequence 11 from Patent WO0248332.
 ACCESSION AX556799
 VERSION AX556799.1 GI:25899907
 KEYWORDS Yersinia pestis
 SOURCE Yersinia pestis
 ORGANISM Yersinia pestis
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Yersinia.
 REFERENCE 1 Short, J., Mathur, E.J., Richardson, T., Robertson, D. and Barton, N.
 AUTHORS Recombinant phytases and uses thereof
 TITLE Patent: WO 0248332-A 11 20-JUN-2002;
 JOURNAL DIVERSA CORPORATION (US)
 FEATURES
 source
 1. .1326
 /organism="Yersinia pestis"
 /mol_type="unassigned DNA"
 /db_xref="taxon:632"

ORIGIN
 Query Match 62.5%; Score 827.6; DB 6; Length 1326;
 Best Local Similarity 77.0%; Pred. No. 6e-237;
 Matches 1021; Conservative 0; Mismatches 304; Indels 1; Gaps 1;
 QY 1 ATGTCAGTATTAGAAATCGGTACGGCTATCTCGGTGTTATGATGCTTAAGCGGATG 60
 Db 1 ATGTCAGTATTAGAAATCGGTACGGCTATCTCGGTGTTATGATGCTTAAGCGGATG 60
 QY 61 GCTGCTATTACTGCGCGGTAGCGCGGAGCCATCGGCTATACCTTAGAACGTTGGTT 120
 Db 61 GCTGCTATTACTGCGCGGTAGCGCGGAGCCATCGGCTATACCTTAGAACGTTGGTT 120
 QY 121 ATTTTGGTCCGATGTTGCTCGCCGACCAAC-ACGCGAGCTTATGATGATGTT 179
 Db 121 ATTTTGGTCCGATGTTGCTCGCCGACCAAC-ACGCGAGCTTATGATGATGTT 180
 QY 180 ACGCCAGATAAGTGGCCACAAATGCGCGGTAAAGCGGGTATTTAAACGCGAGTGGG 239
 Db 181 ACGCCAGATAAGTGGCCACAAATGCGCGGTAAAGCGGGTATTTAAACGCGAGTGGG 240
 QY 240 GAGTTGGTCAATGATGGGGGGTATATGGTGAATTAATTCGACGCTTGGTTGTTA 299
 Db 241 GAGTTGGTCAATGATGGGGGGTATATGGTGAATTAATTCGACGCTTGGTTGTTA 300
 QY 300 GCGCGGGGATGTCGCGCAGAGGGGGGTATATGACAGGAGATATCGATCAACGTACC 359
 Db 301 GCGCGGGGATGTCGCGCAGAGGGGGGTATATGACAGGAGATATCGATCAACGTACC 360

QY 360 CGCTTAACCGGA CAGGCAATTTCTTGATGGTGGCTCCGGGGTGTGGTTGACCGTGCAT 419
 Db 361 CGTTTAAACCGGTACGGCTTTTCTGGATGGTGTGGCGCGGGTTCGGCGCTGACTGTCCAC 420
 QY 420 AATCAGCGCCGATTTGAAAAGACCGATCCCTGTTCCATCCGCTAGAGACTGCGCGTGT 479
 Db 421 AATCAGCGCCGATTTGAAAAGACCGATCCCTGTTCCATCCGCTAGAGACTGCGCGTGT 480
 QY 480 AAGTTAGACACGACCAACAGATTAAGACGATTAAGACGATTTGGCGCGCGCTTTAGAT 539
 Db 481 AAATCGACACGCGCCAAACGATTAAGACGATTTAGAGACGCTTGGCGCGCGCTTTAGAC 540
 QY 540 ACGGTGACGACGCGCTAGCTTAACCTTTTCCAGATGGGACGCTGCTGAATTTTGG 599
 Db 541 ACGGTGACGACGCGCTAGCTTAACCTTTTCCAGATGGGACGCTGCTGAATTTTGG 600
 QY 600 GCTTCTCTTATTCGAATCTTTGCAACAGCAAGAAAACGTTGATTTTGGCCACTTT 659
 Db 601 GCGAGTCGCTACTGCAAGTCACTGACGACGAGGGAACCTTGTGACTTCGCACACTTT 660
 QY 660 GCGGCCATGAGTTACGTTATTAAGAGAGGACCAAGTACCCTCAGTGGGCCACTG 719
 Db 661 GCGGCCATGAGTTATTAAGAGAGGACCAAGTACCCTCAGTGGGCCACTG 720
 QY 720 GCGTTATCATCGACATTTGGGTGAATCTTCTTATTAACAAACGACCAAGCCATGCCAGAG 779
 Db 721 GCGGTGCTAGCAGCTTGGGCGAATCTTCTGCTGAGAACGCGCAGGCGATGCCGAA 780
 QY 780 GTTCCCTGCGNACCGCTAAAGGGCGGAGATTTGGTATCTTATGCTATTACATAAC 839
 Db 781 GTAGCGTGGACGCTTTGAAGGGCGCTGAGAACTGGGTGTCTTCTGAGCCCTGCACAT 840
 QY 840 GCGCAATTTAATTTAATGCGAAACACCGCTATATCCCGCTCATAAAGGAGCGCCATTA 899
 Db 841 GCACAGTTCACTGATGGCTTAACGCCATACATTGACGCCACCAAGGACGCGCTT 900
 QY 900 TTACAGCAGATAGATAGCGCTTTAACCGTTCACTGATGCTGAGGGGCAAGACTACC 959
 Db 901 TTACAGCAATTCGATACCGCACTGACCTGCACTGACGCGCGGCGCAAAACTGCCG 960
 QY 960 ATTTTCAGCCCAAAACCGGCTCTGTTCTCGTGGGCGATGATACCAATATTGCAATATT 1019
 Db 961 ATCTCGGCTCAGAACCGTGTATTCTTCTGGTGGCCACGACCAATATTGCTACATC 1020
 QY 1020 GCGGGTATGCTGGAGCGCATGCGCAGCTACCCGAGCAACTGATTAATCTCCACAGGT 1079
 Db 1021 GCGGGTATGCTGGGCGCAGATTGGCAGTTACCGGAACAACCGGATACACCCACCGGCG 1080
 QY 1080 GGGGATTTGTTTGAATGATGCAAAATCCGGATACCAACGCGCTACGTTGGGTTG 1139
 Db 1081 GCGGCTCTGCTTTGAGCTGTGCGAGATCCGGAATCCGGAATATCACTAACGTTATGTGCGGTT 1140
 QY 1140 AAGATGTTCTACCAACGATGATCAGTTGCGTAAATCCGAAATTTGGATCTGAAAT 1199
 Db 1141 AAGATGTTCTACCAACGATGATCAGTTGCGTAAATCCGGAATTTGGATCTGAAAT 1200
 QY 1200 AACCCGCGGCTATTTTCCGTTGCGATGCTGTTGTTGTAATACGCTGACGATAG 1259
 Db 1201 AATCCGCGGCTATCAGTGTGCTGCTGCGGCTGCGGAGATATGTTGAGATATA 1260
 QY 1260 CTTTCGCGAGCTTGATATCCAAAGAAAGTGGCTAAGGTAATTTGAACCTGCTGCCAC 1319
 Db 1261 CTGTGCGAATCTGATCTTTTCAAAAAAAGTAGGAAAGTCAATTGAACCTGCGTGCAT 1320
 QY 1320 ATCTAA 1325
 Db 1321 ATTTAA 1326

RESULT 6
 AY378096
 LOCUS
 DEFINITION Obesumbacterium proteus phytase gene, complete cds.

ACCESSION AY378096
 VERSION AY378096.1 GI:37266290
 KEYWORDS
 SOURCE Obesumbacterium proteus
 ORGANISM Obesumbacterium proteus
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Obesumbacterium.
 REFERENCE 1 (bases 1 to 2558)
 AUTHORS Zinin,N.V. and Syneokii,S.P.
 JOURNAL Patent: RU (Ru2003127890) -A 17-SEP-2003;
 REFERENCE 2 (bases 1 to 2558)
 AUTHORS Zinin,N.V. and Syneokii,S.P.
 TITLE Revealing and Study of the Characteristics of Bacterial New Phytases
 JOURNAL (in) 2ND MOSCOW INTERNATIONAL CONGRESS "BIOTECHNOLOGY: STATE OF THE ART AND PROSPECTS OF DEVELOPMENT";
 (Nova Science Publishers, Inc., New York, NY, USA)
 REFERENCE 3 (bases 1 to 2558)
 AUTHORS Zinin,N.V., Samsonov,V.V., Borshchevskaya,L.N., Kanikovskaya,A.A., Gudima,M.V., Syneokii,S.P. and Syneokii,S.P.
 TITLE Phytase activity of several bacteria groups
 JOURNAL Biotechnologia (Moscow) 13, 3-10 (2003)
 REFERENCE 4 (bases 1 to 2558)
 AUTHORS Zinin,N.V. and Syneokii,S.P.
 TITLE Direct Submission
 JOURNAL Submitted (02-SEP-2003) VKPM, Gos NII Genetika, 1st Dorozny Proezd, Moscow 117545, Russia
 FEATURES
 Location/Qualifiers
 1..2558
 /organism="Obesumbacterium proteus"
 /mol_type="genomic DNA"
 /db_xref="taxon:82983"
 961..2295
 /note="PhyA"
 /codon_start=1
 /transl_table=11
 /product="phytase"
 /protein_id="AAQ90419.1"
 /db_xref="GI:37266291"
 /translation="MTLSFTHTSPRLKMLPFIASMTTASVASETPSPGYOLE
 KVLISLRHGVAPTKMTQTDVTPNAPWEPVKGVIPTREHLSVLMGGFYKQFQ
 OLGLISKRCPTANDVYVADVDQRTKTEAGFLAPECHELSIHHQDILKQADPLF
 HPVAGKCTMEKTVQQAEOQAGMPIDQLNQHVRPALAMSVINFPKSYCCQHS
 DQCTDLAQMPSKLSIKDNGKVALDGAVALGSLTARI FLLEYAGQMPDAAGHSE
 QDMALLTLHNAQFDLMGRPTFYIAKHNGTPLLQITIVSAINSPSRELPESANKIL
 PFAGHDTNIAINAFMGMSWALPGDPNTPEGALVFERNSDKTKTKITVSQMMYQIL
 AQLNRQETPLTLDRFAGSVALKIPGCDPQTAEGYCPLEPTFTSLAKQNELVCEQ"

Query Match 24.4%; Score 323; DB 1; Length 2558;
 Best Local Similarity 57.8%; Pred. No. 3.1e-85;
 Matches 693; Conservative 0; Mismatches 490; Indels 16; Gaps 6;
 QY 86 CCGAGCCATCGGGCTATCTTTAGAACGCTGTGGTTATTTTTCAGTCGCCATGGTTCGCT 145
 Db 1067 CCGAACCTTCGCGATATCAGCTTTGAAAAGTGGTGATTCTCAGCGCGCATGCGCTTCGAG 1126
 QY 146 CCGCGACCAACA-ACGCGAGCTTATGATGATGTACCGCAGATAAGTGGCCACATCGC 204
 Db 1127 CCGCAACCAAAATGACCAACCATCGCGCAGCTAACACCTAATCGCTGGCTGAGTGGC 1186
 QY 205 CCGTAAACGCGGGTATTTAACGCCACGTCGGTGGCAGTGTGTCTCATTTGATGGGGGGT 264
 Db 1187 CAGTCAAACTGGGGTATATCAGCGCGAGGGCGAGCATTTGGTGAGCGCTGATGGTGGGT 1246
 QY 265 TTTATGCTGATTACTTTTCGAGCCTTGG---TTTGTAGCGCGGATGTCGGGACAGG 321
 Db 1247 TCTATCGGCAAGTTTCAGCAGTTGGGGATCTTGTCTAAAGTTCGCTGTCTCAACGGCGA 1306
 QY 322 GGGGGGTATATGCACAGGCAGATATCGATCAACGCTACCGCTTAAACCGGACAGGCAATTC 381
 Db 1307 ATGATGTTATGTTTGGGACAGCGTCGACCAACGACCTCGGAAACACAGCGAGGCGT 1366

ORIGIN

QY 382 TTGATGGTGGCTCCGGGTGTGGTTTACCGTGCATAATCAGCCGCGATTTGAAAAAGA 441
Db 1367 TAGCGGTCTCGCACTGAAATGCCATCTAAGTATTCATCAGCAAGACATCAGCAGG 1426
QY 442 CCGATCCCTGTTCCATCCGCTAGAGACTGCGGTGTGTAAAGTTAGACAACGCAACAG 501
Db 1427 CCGACCCGCTATTTTCATCCGCTGAAGCGGTGTCTGCACAATGAAAAACCGAGTAC 1486
QY 502 ATAAGCGATTGAAGAACGATTGGCGGGCGGTGTAGATACGGTGAAGCAGCGCTACGCTA 561
Db 1487 AGCAGCGCTGAGCAACAGCGCGGTATGCCATAGATCAGCTAAACAGCAGCTACGCC 1546
QY 562 AACCTTTTCCAGATGGGAGACGCTGCTGAATTTTGGCGCTTCTCCTTATGCAATCTT 621
Db 1547 CTGCACTGGCGCTGATGAGCAGCGTCTGCTGAATTTTCCAAAGTCAACGTACTGCCA 1603
QY 622 TGCAACAGCAGGAAGAACGTTGATTTTGGCCACTTTTGGCGCAATGAAGTTAAAGTTA 681
Db 1604 AACATAGCGCAGCAACCACTGCTGATTTAGCAAGCGCATGCCAAGCAAGCTCTCATCA 1663
QY 682 ATAAGAGGAGCAAAAGTAGTACCCTCAGTGGGCGCACTGGGCTTATCATCAGCAATTTGGGT 741
Db 1664 AAGCAATGGCAATAAAGTGGCACTTGTATGGGCGGTAGTGTCTCAACGCTGGCG 1723
QY 742 AAATCTTCTTATACAAAACGCAACGCAATGCGGAGTGTGCTGCAACGCTAAAG 801
Db 1724 AAATTTCTCTTGAATATGCGCAGGCGATGCCAGATGCGGCTGGGAAAAATCCATT 1783
QY 802 GGGCGGAGAAATGGGTATCTTTATGTCATTACATAACGCGCAATTTAAATTTAATGGCAA 861
Db 1784 CAGAGCAAGACTGGAACGCTTTGCTAACGCTTACAAACGCAATTTGATTTGATGTCAC 1843
QY 862 AAACACCGTATATCGCGCTCATAAAGGAGCGCAATTTATACAGCATAGATACGCTT 921
Db 1844 GCACGCTTATACGCAACATTAACGCGCACCGCTATTACAAACCATGTCAGCGC 1901
QY 922 TAACCTTTCAACTGGATGCTCAGGGGCAAAAGCTACCCATTTCAGGCCAAACCGGGTCT 981
Db 1902 -AATCAATTCAACACCTAGCTCGGAGATTGCGCGGATTCAGCGGATACAAAGTCC 1960
QY 982 TGTTCTCGTGGCGATGATACCAATATGCAATATTTGGGATGCTGCGGAGCGGACT 1041
Db 1961 TGTTCCGCGCGCATGACACCAATATGCAATATCGCGGTATGTTCCGTTATGTCGT 2020
QY 1042 GCGAGCTACCGGAGCAACTGATATCTCCACAGGTGGGGATTTGGTTTGTGAATAT 1101
Db 2021 GGGCGCTACCGCGCAGCGGATTAACAGCTCCAGCGGTGCTTATGTTTGAAGCTT 2080
QY 1102 GGCAAAATCCGGATAACACACGCGCTACGTTGCGGTGAAGATTTCTACCAACGATGG 1161
Db 2081 GGTGAGAT--AAGACGGGGGAAAAATACGTGAGGTACAGATGATGTACCAACGCTGG 2137
QY 1162 ATCAGTTGCGTAATGCCGAAAAATTGGATCTGAAATATCCAGCGGTATTTTTCG 1221
Db 2138 CACAGCTACGTAATCAACCCACTCTCTCTG---GATAGCCCGCTGGCAGCGTGGCG 2194
QY 1222 TTGAGTTGCTGTTGTGAAAAATACCGGTGACGATAAGCTTTGCGAGCTTGATACATTC 1280
Db 2195 TAAAAATACCGGGGTGTGACGATCAACCGCGGAAGGATATGCGCCATAGATACCTTC 2253

RESULT 7
AY425982
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE

AY425982
Obesumbacterium proteus PhylA2 gene, complete cds.
AY425982
AY425982.1 GI:37594476
Obesumbacterium proteus
Obesumbacterium proteus
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Obesumbacterium.
1 (bases 1 to 2161)

AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

Zinin, N.V., Syneky, S.P. and Debabov, V.G.
Obesumbacterium proteus phytase sequence
Patent: RU (2003127890) - A 17-SEP-2003;
VKPM, GosNII Genetika; 1-st Dorozny Proezd, Moscow;
Russia;
2 (bases 1 to 2161)
Zinin, N.V. and Syneky, S.P.
Direct Submission
Submitted (01-OCT-2003) VKPM, GosNII Genetika, 1-st Dorozny Proezd,
Moscow 117545, Russia
Location/Qualifiers
1. 2161
/organism="Obesumbacterium proteus"
/mol_type="genomic DNA"
/db_xref="taxon:82983"
476..1810
/note="phytase; pH 1.5-6.5"
/codon_start=1
/transl_table=11
/product="PhyA2"
/protein_id="AAQ94187.1"
/db_xref="GI:37594477"

CDS

86 CCGAGCCATCGGGCTATCTTTAGAACGTTGGTGTATTTTTCAGTCGCATGTTTCGT 145
582 CGTGCTTCAGCTATCAGCTTGAAGGTGTGATCTCAGCGCTCATGCGTTCGAG 641
146 CCGCGACCAACA-ACGCGCTTATGATGATGTTACGCGAGATGTTGGCCACATGCG 204
642 CCGCGACCAAAATGACGCAACCATGCGGTATGTTTACGCTTAATACCTGGCGTGAATGCG 701
205 CGGTAAACGCGGGTATTTAAACGCGGTTGGGCGGATGTTGCTACATTTAGTGGGGGT 264
702 CTGTTAAGCTGGGTATATCAACCGAGGCGCAATTTAGCGAGCTGATGGGGGAT 761
265 TTTATGTTGATTAATCTTCGACGCTTGG---TTTGTAGCGCGGGATGTTCCGCGAGAG 321
762 TCTATCGACAGAAATTTTCAGCAGTTGGGATCCTGTCTAAAGATCGCTGCCCAACGCGCAA 821
322 GGGGGTATATGACAGCGATATCGATCAACGTACCGCTTACCGCGACGAGCATTTTC 381
822 ATGATGCTATTTTGGGCGAGCGTGCATCAGCGCACCGCGGAAACAGCGCGGGGTTTT 881
382 TTGATGTTGCTGCTCCGGGTGTGTTTGAACGCTGATATTAATCAGCGCGATTTGAAAAAGA 441
882 TAGCGGTCTCGCACCTGAATGCCATCTAAGTATCCATCATCAGCAAGACATCAAGCAGG 941
442 CCGATCCCTGTTCCATCCGCTAGAGACTGCGGTGTGTAAAGTTAGACAACGCAACAG 501
942 TCGATTCGCTATTTTCATCCAGTGAAGCGGTGTCTGCACCATGAAAAAGCGCGGTAC 1001
502 ATAAAGCGATTGAAGAACGATTGGCGGGCGGTGTAGATACGGTGAAGCAGCGCTACGCTA 561
1002 AGCAGCGCTCGAGCAGCGCGGTATGCCATCGCTCAGTTAAACCAACATCTACCGCC 1061
562 AACCTTTTCCAGATGGGAGCGTGTGAATTTTGGGCTTCTCCTTATTTGCAATCTT 621
1062 CTGCACTGGCGCTGATGAGCAACGCTGCTTAATTCGCCACGTCAGCGATTTGCCA---GC 1118
622 TGCAACAGCAGCAAGAAAAACGTTGATTTTTCGCCACTTTTCGCCCAATGAATTAACGTTA 681

ORIGIN

Query Match 24.3%; Score 321.4; DB 1; Length 2161;
Best Local Similarity 57.7%; Pred. No. 9.3e-85;
Matches 692; Conservative 0; Mismatches 491; Indels 16; Gaps 6;
QY 86 CCGAGCCATCGGGCTATCTTTAGAACGTTGGTGTATTTTTCAGTCGCATGTTTCGT 145
Db 582 CGTGCTTCAGCTATCAGCTTGAAGGTGTGATCTCAGCGCTCATGCGTTCGAG 641
QY 146 CCGCGACCAACA-ACGCGCTTATGATGATGTTACGCGAGATGTTGGCCACATGCG 204
Db 642 CCGCGACCAAAATGACGCAACCATGCGGTATGTTTACGCTTAATACCTGGCGTGAATGCG 701
QY 205 CGGTAAACGCGGGTATTTAAACGCGGTTGGGCGGATGTTGCTACATTTAGTGGGGGT 264
Db 702 CTGTTAAGCTGGGTATATCAACCGAGGCGCAATTTAGCGAGCTGATGGGGGAT 761
QY 265 TTTATGTTGATTAATCTTCGACGCTTGG---TTTGTAGCGCGGGATGTTCCGCGAGAG 321
Db 762 TCTATCGACAGAAATTTTCAGCAGTTGGGATCCTGTCTAAAGATCGCTGCCCAACGCGCAA 821
QY 322 GGGGGTATATGACAGCGATATCGATCAACGTACCGCTTACCGCGACGAGCATTTTC 381
Db 822 ATGATGCTATTTTGGGCGAGCGTGCATCAGCGCACCGCGGAAACAGCGCGGGGTTTT 881
QY 382 TTGATGTTGCTGCTCCGGGTGTGTTTGAACGCTGATATTAATCAGCGCGATTTGAAAAAGA 441
Db 882 TAGCGGTCTCGCACCTGAATGCCATCTAAGTATCCATCATCAGCAAGACATCAAGCAGG 941
QY 442 CCGATCCCTGTTCCATCCGCTAGAGACTGCGGTGTGTAAAGTTAGACAACGCAACAG 501
Db 942 TCGATTCGCTATTTTCATCCAGTGAAGCGGTGTCTGCACCATGAAAAAGCGCGGTAC 1001
QY 502 ATAAAGCGATTGAAGAACGATTGGCGGGCGGTGTAGATACGGTGAAGCAGCGCTACGCTA 561
Db 1002 AGCAGCGCTCGAGCAGCGCGGTATGCCATCGCTCAGTTAAACCAACATCTACCGCC 1061
QY 562 AACCTTTTCCAGATGGGAGCGTGTGAATTTTGGGCTTCTCCTTATTTGCAATCTT 621
Db 1062 CTGCACTGGCGCTGATGAGCAACGCTGCTTAATTCGCCACGTCAGCGATTTGCCA---GC 1118
QY 622 TGCAACAGCAGCAAGAAAAACGTTGATTTTTCGCCACTTTTCGCCCAATGAATTAACGTTA 681

```
Db      1119  AGCAGCGCAGACAACTGCACTTTGCGAGAGGTATGCAAGAGCTCTCAATCA 1178
Qy      682  ATAAGAGGAGCAAAAGTGACCTCAGTGGGCCACTGGCGTATATCGACATGGGTG 741
Db      1179  AAGACGATGGCAATAAGTGGGCTCGACGGAGCGGTGGTTGTGCTCAACGCTGGCG 1238
Qy      742  AAATCTTCTATTACAAAACGACAGAGCATCCAGAGGTTCCTGGGAGCGCTAAAG 801
Db      1239  AGATTTCCTACTTGNACATGCTCAGGCGATGCCAAATGCTGCTGGGAAATCCATT 1298
Qy      802  GGGCGAGAAATGGGTATCCTTATTGTCAATACATAACGCGCAATTAATTAATGGCAA 861
Db      1299  CGGAGCAAGACTGGAACTTTACTAACTGCTACAAATGCGCAATTTGATTTGATGTCAC 1358
Qy      862  AAACACCGTATATCGCGCTCATAAAGGACGCCATTATTACAGCAGATAGATACGGCTT 921
Db      1359  GCACGCCCTTACATCGCCAAACATATGGAACCCCGCTATTACAAACCTTCGTTACGCCCA 1418
Qy      922  TAACCTTCAACTGGATGCTCAGGGGCAAAAGCTACCCATTTACGCCCAAAACCGGGTCT 981
Db      1419  TCAACTCAAA--ACTAGCCACAGAAATTGCCAGAGTTATCAGCGGATACAGATCC 1475
Qy      982  TGTTCCTCGTGGGATGATACCAATATTGCCAATATTGGCGGTATGCTGGAGCCGACT 1041
Db      1476  TGTTCCTCGGAGCATGATACCAATATCGCAATATTGCGAGCATGCTCGGCCCTGTGCT 1535
Qy      1042  GGCAGCTACCGAGCAACCTGATACTCTCCACAGGTGGGGGATTTGTTGAACCTAT 1101
Db      1536  GGAGCTACCCGCTCAGCCAGATAACAGCCCTCGGGTGTGCTTTTGAACGTT 1595
Qy      1102  GGCAAATCCGATACACACAGCGCTAGTTCGGTGGTGAAGTGTCTACAAAGATGG 1161
Db      1596  GGTCCGAT--AAGACGGGAAATAATCGTGAGCGCACAGATGATGATCAACAGCTGG 1652
Qy      1162  ATCAGTGGGTGATGCGGAAATTTGGATCTGAAAAATAACCCAGCGGTATTTTCGG 1221
Db      1653  CACAGCTGCGCAATCAACCCACTCACTCTG--GATGAGCTCGGGCGCGTGGT 1709
Qy      1222  TTGCACTGCTGTTGTGAAATACCGTGAGATGAGTTCGAGCTTGATGATCAATTC 1280
Db      1710  TAAAAATACCGGGTGTGACGATCAACCCGCGAGGATTAATTCGCCGCTAGATACCTTC 1768
```

RESULT 8

```
AX042377
LOCUS      AX042377          5421 bp      DNA      linear      PAT 23-NOV-2000
DEFINITION Sequence 6 from Patent WO0064247.
ACCESSION  AX042377
VERSION     AX042377.1  GI:11340995
```

```
KEYWORDS   synthetic construct
SOURCE      synthetic construct
            artificial sequences.
```

REFERENCE 1

```
AUTHORS    Forberg, C.W., Golovan, S. and Phillips, J.P.
TITLE       Transgenic animals expressing salivary proteins
JOURNAL     Patent: WO 0064247-A 6 02-NOV-2000;
            UNIVERSITY OF GUELPH (CA)
```

FEATURES

```
source      Location/Qualifiers
            1..5421
            /organism="synthetic construct"
            /mol_type="unassigned DNA"
            /db_xref="taxon:32630"
            /note="SV40/APPA + intron plasmid"
```

ORIGIN

```
Query Match      17.1%; Score 226.4; DB 6; Length 5421;
Best Local Similarity 53.3%; Pred. No. 3.5e-56;
Matches 572; Conservative 0; Mismatches 487; Indels 14; Gaps 4;
```

```
Qy      107  TAGACGCTGGTGTATTATTTGATCGCCATGCTGTTCCGCTCGCCGACCAA-ACACGCGACG 165
```

```
Db      128  TGAAGAGTGTGGTGAATTGTCACTGCTCATGTTGTGCTGCTCCAAACCAAGGCCACGCAAC 187
Qy      166  TTATGATGATGTTATGCCAGATAAGTGGCCA CAATGGCGCGTAAAGCGGGGTATTTAA 225
Db      188  TGATGCAAGATGTCACCCAGACGATGGCCAACTGGCGGTAAAACTGGGTGGCTGA 247
Qy      226  CGCACGCTGTCGGGAGTTGTCACATTGATGGGGGGTTTTATGTGTGATTACTTTCCGCA 285
Db      248  CACCGCGNGGTGGTGAAGCTAATCGCTAATCTCGGACATTAACCAACGCCAGCGTCTGGTAG 307
Qy      286  GCCTTGTGTTTGTAGCG--GCGGATGTCCGSCAGAGGGGGGGGTATATGCAACAGCAG 342
Db      308  CCGACGGAATGCTGGCGAAAAAGGGCTGCCCGCAGTCTGTCAGGTCCGGATTAATTGCTG 367
Qy      343  ATATCGATCAACGTCACCGCTTAACCGGACAGCAATTTCTTGATGTGTGGCTCCGCGGT 402
Db      368  ATGTGACGAGCGTACCCGTAAAACAGGCGAAGCTTTGCCCGCCGGGTGGCACTGACT 427
Qy      403  GTGTTTGAACCGTGCATPAATCAGGCCGATTTGAAAAGACCGATCCCTGTTCATTCGCG 462
Db      428  GTGCAATAACCGTACATACCCAGCGAGATACGTCAGTCCGATCCGATTAATTAAATCCTC 487
Qy      463  TAGAGACTGCGGTGTGTAAGTTAGACAACGCAACACAGATAAGCGATTAAGAACGAT 522
Db      488  TAAAACTGGCGTTTGCCTACTGATACGGAACGTGACTGACCGGATCCTCAGCAGGG 547
Qy      523  TGGCGCGCGCTTAGATACGCTGAGCCAGCGCTACGCTAAACCTTTTGCCAGATCGGG 582
Db      548  CAGAGGGTCAAATTGCTGACTTTACCGGGCATCGGCAACGGCGTTTCGCGAATCGAAC 607
Qy      583  ACGTGTGATTTTGGGCTTCTCTTATGCAATCTTTGCAACAGCAGGAAACCT 642
Db      608  GGGTGTCTTAATTTTCGCAATCAAACTTGTGCTTAAACGTGAGAAACAGGACGAAGCT 667
Qy      643  GTGATTTTGGCCACTTTGCGGCCAATGAAGTTAAACGTTAATAAGAGGACAAAAAGTGA 702
Db      668  GTTCATTAAACGAGGCATTACCATCGGAATCAAGGTGAGGCGCGA-----CAATGTCT 721
Qy      703  CCCTCAGTGGCCACTGGGCTTATCATCGACATGGGTGTAATCTTTATACAAAACG 762
Db      722  CATTAACCGGTGGGTAAAGCTCGCATCAATGCTGACGAGATATTTCCTCTGCAACAAG 781
Qy      763  CACAAGCATGCGCAGAGTTGCTCGCAACGGCTAAAGGGCGGAGAAATGGGTATCCT 822
Db      782  CACAGGGAATGCCGAGCGCGGGTGGGAGGATCACCGATTCAACACCGAGTGAACACCT 841
Qy      823  TATTGTCAATTAACAGCGCAATTTAATTAATGCAAAAACACCGTATATCGCCGCTC 882
Db      842  TGCTAAGTTTGCATACGCGCAATTTTATTGTCTACAACGCGCAGAGGTTGCCCGCA 901
Qy      883  ATAAAGGACGCGCATTTATTACAGCAGATAGATACGGCTTTAACCTTTCACTGGATGCTC 942
Db      902  GCGCGCCACCCGTTATTAGATTGATCAAGACAGGGTTGACGCCCCACCGCAAAA 961
Qy      943  AGGGGCAAAAGCTACCCATTTACGCCCAAAACCGGGTCTTTGCTCGGTGGGATGATA 1002
Db      962  ACAGCGTATGTTGTGACATTAACCACTTCA-----GTGCTGTTTATCGCGGACAGATA 1017
Qy      1003  CCAATATTGCCAATATTGCGGGTATGCTGGAGCGCGACTGGCAGCTACCGAGCAACCTG 1062
Db      1018  CTAATCTGGCAATCTCGCGCGCACTGGAGCTCACTGAGCTTCCCGGTGACCGCG 1077
Qy      1063  ATAACTCTCCACGAGTGGGGATTTGGTTTTGAACTATGGCAAAATCCGGATAACCAAC 1122
Db      1078  ATAAACGCGCGCAGGTGGTGAATGGTGTGTAACGCTGCGCTCGGCTAAGCGATAACA 1137
Qy      1123  AGCGCTACGTTGGGTGAAGATGTTCTACCAACGATGGATCAGTTGCGGTAAT 1175
Db      1138  GCCAGTGGATTCAGGTTTCGCTCTCCAGACTTTACAGCAGATCGGTGAT 1190
```

```
RESULT 9
ECOACPHD
```

LOCUS ECOACPHD 1901 bp DNA linear BCT 04-APR-2002
 DEFINITION Escherichia coli periplasmic phosphoanhydride phosphohydrolase
 (appa) gene, complete cds.
 ACCESSION L03373.1 GI:145182
 VERSION L03373.1
 KEYWORDS acid phosphatase; periplasmic phosphoanhydride phosphohydrolase.
 SOURCE Escherichia coli
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Escherichia.
 REFERENCE 1 (sites)
 Dassa, J., Marck, C. and Boquet, P.L.
 The complete nucleotide sequence of the Escherichia coli gene appA
 reveals significant homology between pH 2.5 acid phosphatase and
 glucose-1-phosphatase
 J. Bacteriol. 172 (9), 5497-5500 (1990)
 JOURNAL 90368616
 MEDLINE 2168385
 PUBMED
 REFERENCE 2 (bases 1 to 1901)
 Ostranin, K., Harms, E.H., Stevis, P.E., Kuciel, R., Zhou, M.M. and Van
 Eeten, R.L.
 Overexpression, site-directed mutagenesis, and mechanism of
 Escherichia coli acid phosphatase
 J. Biol. Chem. 267 (32), 22830-22836 (1992)
 JOURNAL 93054596
 MEDLINE
 PUBMED 1429631
 FEATURES
 source
 /organism="Escherichia coli"
 /mol_type="genomic DNA"
 /db_xref="taxon:562"
 gene
 188..1486
 /gene="appa"
 CDS
 188..1486
 /gene="appa"
 /standard_name="acid phosphatase"
 /citation=[1]
 /codon_start=1
 /transl_table=11
 /product="periplasmic phosphoanhydride phosphohydrolase"
 /protein_id="AAA0005.1"
 /db_xref="GI:145182"
 /translation="MKAILIPFLSLLIPLPQAPQSEBELKESVIVSRHGVAP
 TKATLMQDVTTPAWPTVPVKLGLNLTFRGGLIAYLGHYQARLVADGLKAGKCPQS
 GOVAIADVPTKRTGEAFAGLADPCATVHTQADTSPDFLPLNPLKTVGQCLDNA
 NVTDAILDSVSGTADTGHRTAFRLERVLNFPQSNCLKREKQDESCLTQALPS
 ELKVSANSLTGAVSLASMLTEIFLQQAQGMPECGSITDISHQWNTLLSLHNAQF
 YLLORTVEARSATPLDLIKTALTPHPQKAYGVLTPTSVLFIAGHDNLNLAGG
 ALELNWLPQPDNTPGGSLVFERWRLNSQWIOVSLVFQTLQMRDKTFLSLNT
 PPGEVKTLTAGCEERNAQGMCSLAGFTQIVNEARI PACSL"
 variation
 299..301
 /gene="appa"
 /citation=[1]
 /phenotype="R16A mutant"
 ORIGIN
 Query Match 17.0%; Score 225.2; DB 1; Length 1901;
 Best Local Similarity 53.2%; Pred. No. 7.7e-56;
 Matches 571; Conservative 0; Mismatches 489; Indels 13; Gaps 4;
 QY 107 TAGAACGTGTGTTATTTGATGCGCATGTGTTGCTGCGCACCAAA-ACACGCGAGC 165
 Db TGGAAAGTGTGTTATTTGATGCGCATGTGTTGCTGCGCACCAAA-ACACGCGAGC 165
 QY 166 TTATGAATGATGTTACGCCAGATAGTGGCCCAATGCGCGTAAAAAGCGGGGTATTAA 225
 Db TGTATGAGGATGTACCCCGACGATGGCCCAATGCGCGTAAAAAGCGGGGTATTAA 225
 QY 226 CGCCACGTGTGCGGAGTGTGTCATGATGCGGGGGGTTTATGTTGATTTACTTCGCA 285
 Db CACCGCGGGGTGGTGGAGTATCGCTATCTCGGACATTACCAAGCCGCGTGTGTAG 455
 QY 286 GCCTTGTTGTTTACG---CGGGGATTCGCGGAGAGGGGGGGGTATATGACAGGCGAG 342

456 CCGACGGATGTGTCGCAAAAAGGGCTCCCGCAGTCTGGTCAGGTCCGATTATTGCTG 515
 QY 343 ATATCGATCAACGTACTCCGCTTAACCGGACAGGCAATTTCTTGATGGTGTGCTCGGGGT 402
 Db 516 ATGTGACGAGGTACCGGTAAACACAGCGGAGCCTTCGCGCGGGGTGCACTGACT 575
 QY 403 GTGGTTTACCGGTGATTAATCAGGCGGATTTGAAAAGACCGATCCCTCTTTCATCCGG 462
 Db 576 GTGCAATAACCGTATACCCAGGCGAGATAGTCCAGTCCCGATCCGTTATTTATCTC 635
 QY 463 TAGAGACTGGCGTGTAAAGTTAGACAACGACACAGATAAAGCGATTGAAGACGAT 522
 Db 636 TAAAACCTGGCGTTTCCGACTGGATACGGAAGTCACTGACCGATCTCAGCAGG 695
 QY 523 TGGGCGGCGGTAGATACGTTGAGCCAGCGTACGCTAAACCTTTTCCGAGATGGGG 582
 Db 696 CAGGAGGGTCAATTGCTGACTTTACCGGCGCATCGCAACCGCGTTTCCGCACTGGAAC 755
 QY 583 AGTCTGTAATTTTCGGCTCTCTCTTATTCMAATCTTTCACACAGCAAGAAAACGT 642
 Db 756 GGGTGTCTAATTTTCCGCAATCAAACTGTGCTTAAACGTGAGAAACAGGACGAAAGCT 815
 QY 643 GTGATTTTGGCCACTTTGCGGCCAATGAAGTTAAGCTTAATAAAGAGGGGACAAAGTGA 702
 Db 816 GTTCATTAAGCGCAGCATTTACCATCGAACTCAAGGTGAGCGCGA-----CAATGTCT 869
 QY 703 CCTCAGTGGGCGCACTGGCGTTATCATCGACATTTGGGTGAAATCTTCTATTACAAAAG 762
 Db 870 CATTAAACCGGTGCGTAAAGCTTCGCATCAATGCTGACGAGATATTTCTCTGCAACAG 929
 QY 763 CACAAGCCATGCGCAGAGGTTCCTGCGCAACCGCTTAAAGGGCGGAGAAATTTGGGTATCCT 822
 Db 930 CACAGGGAATGCGGAGCGCGGGTGGGAAGATCAACCGATTTCACACAGTGAACACCT 989
 QY 823 TATTGTCTATTACAAACGGGCAATTTAATTAATGCGAAAACACCGGTATATCGCCGCTC 882
 Db 990 TGTAAAGTTTGCAATACCGCAATTTTATTTGCTCAACGACGCGCAGGTTGCGCGCA 1049
 QY 883 ATAAAGGAGCGGCATTATTACAGCAGATAGATACGGCTTAAACCTTCAACTGTGATGCTC 942
 Db 1050 GCGCGCCACCCCGTTATTAGATTGATCAAGACAGCGTTGACGCCCCATC---CACCGC 1106
 QY 943 AGGGGCAAAAGTACCCATTTTACGCCCAAAACCGGGTCTTGTCTCGTGGGCGATGATA 1002
 Db 1107 AAAAAACAGGCGTATGGTGTGACATTTACCCACTTTCAGTGTCTGTTTATCGCCGACACGATA 1166
 QY 1003 CCAATATTGCCAATATTCGGGTATGCTGGAGCGGACTGGCAGCTACCCGAGCAACCTG 1062
 Db 1167 CTAACTGGCAATCTCGCGCGGCACCTGGAGCTCACTGGACGTTCCCGTCAAGCGG 1226
 QY 1063 ATAATACTCCACCGGTGGGGAATGGTTTGTAACTATGCAAAATCCGGATAACACC 1122
 Db 1227 ATAACAGCGCGCGCAGGTGTGAACTGTGTTTGAACGCTGGCGTCCGCTAAGCGATAACA 1286
 QY 1123 ACGGTACGTTGCGGTGAGATGTTCTACCAACGATGATGATGATGATGATGAT 1175
 Db 1287 GCCAGTGGATTCAGGTTTGGTGGTCTTCCAGACTTTTACAGCAGATGCGGTAT 1339
 RESULT 10
 ECOAPPA
 LOCUS
 DEFINITION Escherichia coli periplasmic phosphoanhydride phosphohydrolase
 (appa) gene, complete cds.
 ACCESSION M58708
 VERSION M58708.1 GI:145283
 KEYWORDS
 SOURCE Escherichia coli
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Escherichia.
 REFERENCE 1 (bases 1 to 1901)

AUTHORS Dassa,J., Marck,C. and Boquet,P.L.
TITLE The complete nucleotide sequence of the Escherichia coli gene appA reveals significant homology between pH 2.5 acid phosphatase and glucose-1-phosphatase
J. Bacteriol. 172 (9), 5497-5500 (1990)
JOURNAL MEDLINE 90368616
PUBMED 2169385
REFERENCE 2 (sites)
AUTHORS Ostanin,K., Harms,E.H., Stevis,P.E., Kuciel,R., Zhou,M.M. and Van Etten,R.L.
TITLE Overexpression, site-directed mutagenesis, and mechanism of Escherichia coli acid phosphatase
J. Biol. Chem. 267 (32), 22830-22836 (1992)
JOURNAL MEDLINE 93054596
PUBMED 1429631
REFERENCE 3 (sites)
AUTHORS Ostanin,K. and Van Etten,R.L.
TITLE Asp304 of Escherichia coli acid phosphatase is involved in leaving group protonation
J. Biol. Chem. 268 (28), 20778-20784 (1993)
JOURNAL MEDLINE 94012612
PUBMED 8407904
FEATURES source
Location/Qualifiers
1..1901
/organism="Escherichia coli"
/mol_type="genomic DNA"
/db_xref="taxon:562"
3..1522
/gene="appA"
3..8
/gene="appA"
16..108
/gene="appA"
/notes="ORF"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AA72085.1"
/db_xref="GI:145284"
/translation="MWLLWFGILLMCSLSLTLVWLDPLRLKs"
118..149
/gene="appA"
133..138
/gene="appA"
156..161
/gene="appA"
172..177
/gene="appA"
188..1486
/gene="appA"
/standard_name="acid phosphatase"
/codon_start=1
/transl_table=11
/product="periplasmic phosphoanhydride phosphohydrolase"
/protein_id="AA72086.1"
/db_xref="GI:145285"
translations="MKAILIFLSLILPTPOSFAQSEPELKLESVIVSRHGVRA
TKATLMQDVTDPATWTPVKLGWLTFRGELIAYLGHYQRQLVADGLLAKKGCPOS
GOVAIIADVERGRTGEAPAGLAPDCAITVTOADTSPPDLPFNPLKTCQLDNA
NVTDAILSRAGGSIAIDTGHROTAFRELRVLPFOSNCLKREKODESCSLTOALPS
ELKVSADNLSLTAVSLASMLTEIFLLQQAQMPFEGWGRITDSDHWNLLSLRNAQF
YLQRTPEVARSATLLELDLIKALTPHPFQGAQVTVLFTVLFAGHDTMLNLGG
ALELNTLPQPPNTPPGGELVFRWRRLSDNSQWIOVSLVFQTLQOQMDKTPLSLNT
PPGEVLTLAGCBERNAAQMGCSLAGFTQIVNEARIIPACSL"
188..253
/gene="appA"
254..1483
/gene="appA"
/product="periplasmic phosphoanhydride phosphohydrolase"
299..301
/gene="appA"
/standard_name="H303A mutant"
/note="created by site directed mutagenesis"

/citation=[2]
299..301
/gene="appA"
/standard_name="R16A mutant"
/notes="created by site directed mutagenesis"
/citation=[2]
299..301
/gene="appA"
/standard_name="R20A mutant"
/notes="created by site directed mutagenesis"
/citation=[2]
299..301
/gene="appA"
/standard_name="R92A mutant"
/notes="created by site directed mutagenesis"
/citation=[2]
302..304
/gene="appA"
/standard_name="H17N mutant"
/notes="created by site directed mutagenesis"
/citation=[2]
1163..1165
/gene="appA"
/standard_name="D304A mutant"
/notes="created by site directed mutagenesis"
/citation=[3]
/phenotype="Asp 304 changed to Ala"
/replace="gct"
1163..1165
/gene="appA"
/standard_name="D304E mutant"
/notes="created by site directed mutagenesis"
/citation=[3]
/phenotype="Asp 304 changed to Glu"
/replace="gag"
1497..1522
/gene="appA"
/note="rho-independent terminator"
ORIGIN
Query Match 17.08; Score 225.2; DB 1; Length 1901;
Best Local Similarity 53.2%; Pred. No. 7.7e-56;
Matches 571; Conservative 0; Mismatches 489; Indels 13; Gaps 4;
QY 107 TAGAACGTGTGGTTATTTTGAGTCGCCATCGTGTCTCGCCGACCAA-ACAACGAGC 165
DB 276 TGGAAAGTGTGTGATTGTGTCATGTCGTGTGCTGTCCAAACCAAGCCACGCAAC 335
QY 166 TTATGAATGATGTTACCGCAGATAAGTGGCCACAATGCGCGGTAAGGCGGGTATTAA 225
DB 336 TGAATGCAGGATGTCAACCCAGACGCATGCGCCGTAACCTGCGCGTAAACTGGTGTGCTGA 395
QY 226 CGCCACGTGTGCGGAGTTGTCACATTGATGGGGGGTTTTATGGTGATTACTTTCCGA 285
DB 396 CACCGCGNGTGTGTGAGCTAATCGCCTATCTCGACATTACCAACGCGCGTCTGGTAG 455
QY 286 GCCTTGGTTTGTAGCG---CGCGGATGTCCGCGAGAGGGGGGTATATGCACAGCAG 342
DB 456 CCGACGGATTCTCGCGGAAAAAGGGCTGCCCGCAGTCTGTGTGAGTTCGGATTTTGTCTG 515
QY 343 ATATCGATCAACGTACCCGCTTAACCGCAGACAGGCAATTTCTTGATGGTGTGGCTCCGGGGT 402
DB 516 ATGTGCAGACGCTACCCGTAAACACAGCGGAGAGCTTCGCGCGCGGGTGGCACTGACT 575
QY 403 GTGGTTTGACCGTCATATCAGCCCGATTGAAAAAGACGATCCCTGTTCATCCGG 462
DB 576 GTGCAATAACCGTACATACCCAGGCAGATACGTCCCGATCCGTTATTTAATCTC 635
QY 463 TAGAGACTGGCGTGTGTAAGTTAGACACGACACACAGATAAAGCCGATTGAAGAACGAT 522
DB 636 TAAAAACTGGCGTTTGGCCACTGGATAACGCGACGCTGACTGACGCGATCTCTCAGCAGG 695
QY 523 TGGCGCGCGCTTAGATACGGTGAGCCGCGCTACGCTAAACCTTTTGGCCCGAGATGGGG 582

Db 696 CAGAGGGTCAATGCTGACATTTACCGGCGATCGGCAACCGGCTTTCGGGAAGTGAAC 755
Qy 583 ACGTGCTGAATTTTCGGGCTTCTCTTATTTGCAAAATCTTTTGCAACAGCAAGGAAACGCT 642
Db 756 GGGTGTCTAATTTTCGGCAATCAAACTTGTGCTTAAACGCTGAGAAACAGACGAAAGCT 815
Qy 643 GTGATTTTGGCCACTTTGGCGGCAATGAAGTTAAAGTTAAAGTAAAGGAGGACAAAAGTGA 702
Db 816 GTTCATTAACGCGAGGCAATACCAATCGGAATCTCAAGGTGAGCGCGA-----CAATGTCT 869
Qy 703 CCCTCAGTGGCGCACTGGGCTTATCATCGCAATTTGGGTGAAATCTCTTTATTAACAACAG 762
Db 870 CATTAACCGGTGCGGTAAAGCTCGCATCAATGCTGACGGAGATTTCTCTCTGCAACAG 929
Qy 763 CACAGCCATGCGAGGTTGCTGCGCAACGCTTAAAGGGGGGAGAAATGGGTATCCT 822
Db 930 CACAGGGAATGCGCGGAGGTTGCGGAGGAGGATCACCGAATCACACGAGTGAACACCT 989
Qy 823 TATTGTCAATACATAACGCGCAATTTAAATTTAATGCGCAAAACACCGTATATCGCCGCTC 882
Db 990 TGCTAAGTTGCAATACGCGCAATTTTATTTGCTACACGCAACGCGAGAGTTGCCGCA 1049
Qy 883 ATAAAGGAGCGCCATTAATACAGCAGATAGATAGCGGCTTTAAACCTTTCAACTGATGCTC 942
Db 1050 GCCCGCGCACCCCGCTTATTAGATTGATCAAGACAGCGGTTGACGCCCATC---CACCGC 1106
Qy 943 AGGCGCAAAAGCTACCCATTTACGCCCAAAACCGGCTTGTCTCTCGGTGGGATGATA 1002
Db 1107 AAAAAGCGGTATGCTGAGATTTTATTTGCTACACGCAACGCGAGAGTTGCCGCA 1166
Qy 1003 CCAATATTGCAATATTTCGGGTATGCTGGAGCGGAGTGGAGTACCGAGTACCGAGCAACCTG 1062
Db 1167 CTAATCTGGCAATCTCGCGCGGCACTGGAGCTCAACTGGAGCTCAACTGGAGCTTCCCGGTCAGCCGG 1226
Qy 1063 ATATATCTCCACAGGTGGGGATTTGGTTTTTGAATTAATGCGCAAAACACCGTATATCGCCGCTC 1122
Db 1227 ATAAACCGCCCGAGGTGGTGAATTTGTTTGAACGCTGGCTGCGGTGAGCGATGAACA 1286
Qy 1123 AGCGCTACGTTGCGGTGAAGATGTTCTACCAACGATGGATCAGTTCGGTAAT 1175
Db 1287 GCCAGTGGATTGAGTTTCGCTGCTCTCCAGACTTTTACAGCAGATGCGTGAT 1339

RESULT 11
AX356570 1901 bp DNA linear PAT 06-FEB-2002
LOCUS AX356570 Sequence 5 from Patent WO0190333.
DEFINITION AX356570
ACCESSION AX356570
VERSION AX356570.1 GI:18621057
KEYWORDS Escherichia coli
SOURCE Escherichia coli
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1 Short,J.M., Kretz,K.A., Gray,K.A., Barton,N.R., Garrett,J.B. and O'Donoghue,E.
AUTHORS Recombinant bacterial phytases and uses thereof
TITLE Patent: WO 0190333-A 5 29-NOV-2001;
JOURNAL DIVERSA CORPORATION (US)
FEATURES Location/Qualifiers
source 1..1901
/organism="Escherichia coli"
/no_type="unassigned DNA"
/db_xref="taxon:562"
ORIGIN
Query Match 17.0%; Score 225.2; DB 6; Length 1901;
Best Local Similarity 53.2%; Pred. No. 7.7e-56;
Matches 571; Conservative 0; Mismatches 489; Indels 13; Gaps 4;
Qy 107 TAGAAGCTGTGGTTATTTGAGTCGCCATGGTGTTCGCTCGCGCAACAA-ACACGCGAGC 165

Db 276 TGGAAAGTGTGGTGTGATTTGCTAGTGTGATGTTGCTCTCAACCCAGGCCACGCAAC 335
Qy 166 TTATGAATGATGTGTACGCCAGATAGTGGCCACAAATGCGCGGTAAAGCGGGGTATTATAA 225
Db 336 TGATGCAAGATGTCAACCCAGACGATGCGCAACCTTGGCCGGTAAAACTGGGTGTGACTGA 395
Qy 226 CGCCACGCTGGTGGGAGTTGCTCAATTTGATGGGGGGTTTTATGTTGATTACTTTTCGCA 285
Db 396 CACCCGCGGTGGTGGAGCTAATTCGCTATCTCGACATACCAACGCCAGCGCTGTGGTAG 455
Qy 286 GCCTTGGTTTGTATAGCG--CGCGGATGTCGGGCAAGAGGGGGGGGTATATGCAACAGGAG 342
Db 456 CCGACGATTTGTCGCGGAAAAAGGGCTGCCCGAGTCTGGTCAGGTCGCGGATTTTGTCTG 515
Qy 343 ATATCGATCAACGTACCGCTTTAACCGGACAGGCAATTTCTTGTGCTGTGGTCTCGGGGT 402
Db 516 ATGTGCAAGAGGCTTACCCGTAAACAGGCGAGCTTTCGCCCGGGGTGCACTGACT 575
Qy 403 GTGGTTTTCACCGTGCATTAATCAGGCGGATTTGAAAAAGACCGATCCCTGTTCATCCGG 462
Db 576 GTGCAATAACCGTACATACCCAGGCGAGATAGTCCAGTCCCGATCCGTTATTTAATCCTC 635
Qy 463 TAGAGACTGGCGGTGTAAAGTTAGTACAGCAGCAGACAGATAAAGCGGATTTGAAGAACAT 522
Db 636 TAAAAACTGGCGTTTTCGCAACTGGGATAACGCGAAGCTGACTGACCGGATCTCTCAGCAGG 695
Qy 523 TGGCGGGCGGCTTAGATACGGTGGAGCGCTAGCTTAAACCTTTTCCCGAGTGGGG 582
Db 696 CAGGAGGGTCAATTTGCTGACTTTTACGGGCAATCGGCAAAACGGCGTTTCGCGAAGTGAAC 755
Qy 583 AGTGTGTGAATTTTGGGCTTCTCTTATTTGCAAAATCTTTGCAACAGCAGCAAGAAAAAGCT 642
Db 756 GGGTGTCTTAATTTTCGCAATCAAACTTTGTGCTTTAAACGTGAGAAACAGGACGAAAGCT 815
Qy 643 GTGATTTTCCCACTTTGGGGCAATGAAGTTAAAGTTAAATAAAGAGGGAACAAAGTGA 702
Db 816 GTTCATTAACGAGGCAATTTACATCGGAATCAAGTGTAGCGCGA-----CAATGTCT 869
Qy 703 CCCTCAGTGGGCACTGGGCTTATCATCGACATTTGGGTGAATCTCTTTATTAACAACG 762
Db 870 CATTAACCGGTGCGGTAAAGCTCGCATCAATGCTCAGCGAGATATTTCTCTGCAACAG 929
Qy 763 CACAGCCATGCGAGGTTGCTTGGCAACGCTTAAAGGGGCGGAGAAATTCGGTATCCT 822
Db 930 CACAGGGAATGCGGAGCGGCGGTGGGAGAGATCACCGATTCAACCGATGGAAACACCT 989
Qy 823 TATTGTCAATTAATAACGCGCAATTTAAATTTAATGGCAAAACACCGTATATTCGCCGCTC 882
Db 990 TGCTAAGTTTGCATAACGCGCAATTTTATTTGCTACACGCAACGCGAGAGTTGCCGCA 1049
Qy 883 ATAAAGGAGCGCATTTATTAACAGCAGATAGATACGCTTTTACCTTTCACTGGATGCTC 942
Db 1050 GCGCGCCACCCCGTTATTAGATTTGATCAAGACAGCTTTGAGCCCATC---CACCGC 1106
Qy 943 AGGCGCAAAAGCTACCCATTTACGCCCAAAACCGGCTTGTTCCTCGTGGGCAATGATA 1002
Db 1107 AAAAAGCGGTATGCTGAGATTTTACATTAACCCACTTTCAGTGTCTTATTCGCGGACAGATA 1166
Qy 1003 CCAATATTGCAATATTTCGGGTATGCTGGAGCGGAGTGGAGTACCGAGTACCGAGCAACCTG 1062
Db 1167 CTAATCTGGCAATCTCGCGCGGCACTGGAGCTCAACTGGAGCTTCCCGGTCAGCCGG 1226
Qy 1063 ATAACTACTCCACAGGTGGGGATTTGGTTTTTGAATTAATGCGCAAAACACCGTATTAACCC 1122
Db 1227 ATAAACGCGCGCGAGGTGGTGAATTTGTTTGAACGCTGGCTGCGGTGAGCGATGAACA 1286
Qy 1123 AGCGCTACGTTGCGGTGAAGATGTTCTACCAACGATGGATCAGTTCGGTAAT 1175
Db 1287 GCCAGTGGATTGAGTTTCGCTGCTCTCCAGACTTTTACAGCAGATGCGTGAT 1339

930	db	CA CAGGGAATCCGGAGCCGGGTGGGGAAGGATCA	CGGATTCACACCAAGTGGAAACACT	989
823	yy	TATTGTCAATTACATAACGCGCAATTAATTAAT	TATGCGCAAAAAACACCGGTATATCCCGGTC	882
990	db	TGCTAAGTTTGCATAACGCGCAATTTATTTT	TGCTTACAACGACGCGCAGAGTTGCCCGCA	1049
883	yy	ATAAAGGAGCCCATTAATACAGCAGATAGAT	CAGGCTTTAAACCTTCAACTGGATGCTC	942
1050	db	GCCGCGCCACCCCGTTATTAGATTGATCAAG	ACAGCGTTGACGCCCCCATC--CACGCG	1106
943	yy	AGGGGCAAAAGCTACCCATTTACGCCCAAAAC	CGGGTCTTGTTCCTCGGTGGGCATGATA	1002
1107	db	AAAAACAGGCGTATGGTGTGCACATTACCAC	CTTCAGTGTGTTTATCGCGGACACGATA	1166
1003	yy	CGAATATTGCCAATATTGCGGGTATGCTGGG	AGCCGACTGCGCAGTACCCGAGCAACTG	1082
1167	db	CTAATCTGGCAATCTCGCGGCGCACTGGAG	CTCAACTGGACGCTTCCCGGTGAGCGCG	1226
1063	yy	ATAATACTCCACAGGTGGGGGATTGGTTTT	TTCGAATCATGCGCAAAATCCGATAACACC	1122
1227	db	ATACACGCCGCCAGTGGTGACTGGTGT	TGAAACGCTGGCGTGAAGCGATAACA	1286
1123	yy	AGCGTACGTTGGGTGAAAGATGTTCTACAA	AGATGGATCAGTTGCGTAAAT	1175
1287	db	GCCAGTGGATTACAGTTTCGTGCTCTCCAG	ATTTCCAGCAGATCGTGAT	1339

RESULT 13

AX356572

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AX356572

Sequence 7 from Patent WO0190333.

AX356572.1

GI:18621059

Escherichia coli

Escherichia coli

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.

1

Short, J.M., Kretz, K.A., Gray, K.A., Barton, N.R., Garrett, J.B. and O'Donoghue, E.

Recombinant bacterial phytases and uses thereof

Patent: WO 0190333-A 7 29-NOV-2001;

DIVERSA CORPORATION (US)

Location/Qualifiers

1. .1901

/organism="Escherichia coli"

/mol_type="unassigned DNA"

/db_xref="taxon:562"

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1

Short, J.M., Kretz, K.A., Gray, K.A., Barton, N.R., Garrett, J.B. and O'Donoghue, E.

Recombinant bacterial phytases and uses thereof

Patent: WO 0190333-A 7 29-NOV-2001;

DIVERSA CORPORATION (US)

Location/Qualifiers

1. .1901

/organism="Escherichia coli"

/mol_type="unassigned DNA"

/db_xref="taxon:562"

ORIGIN

Query Match

Best Local Similarity

Matches

571; Conservative

0; Mismatches

489; Indels

13; Gaps

4;

107

TAGAACGTGTGGTTATTTTCAGTCGCCATGGTGTTCGCTCGCCGACCAA-ACAAAGCAGC

165

276

TGGAAGATGGTGATTTGTCAGTCGTCATGCTGTCGTGTCGTCCAAACGAGCCACGCAAC

335

166

TTATGAATGATGTTTACGCCAGATAAGTGGGCCACAATGCGCGGTAAAAAGCGGGGTATTTAA

225

336

TGATGCAGGATGTCACCCAGACGCATGGCCAACTGCGCGTAAACTGGTGGCTGA

395

226

CGCCACGTGTGCGGAGTTGGTCAATGTATGGGGGGTTTTATGGTGATTACTTTCCCA

285

396

CACCGCGNGTGGTGAGCTAATCGCCTATCTCGGACATTACCAACGCCAGCGTCTGGTAG

455

286

GCCTTGGTTTGTATAGCG---GCGGGATGTCCGCGAGAGGGGGGGGTATATGCACAGGCAG

342

456

CCGACGGAATGCTGGCGAAAAAGGCTGCGCGCAGTCTGGTCAGTTCGGATTATTGCTG

515

343

ATATTCGATCAACGATACCCGCTTAACCGGACAGGCATTTCTTGATGGTGTGGCTCGGGGT

402

Db 516 ATGTGCGACGAGGTACCGGTAAACACAGCGAGAGCCTTGC CGCGCGGCTGGCACCTGACT 575
Qy 403 GTGTTTACCGTGCATATCAGCGCGATTTTGAAGAACAGCGATCCCTCTGTTCATCCGG 462
Db 576 GTGCAATAACCGTACATACCGAGGAGATACGTCCAGTCCCGATCCGTATTATATCCTC 635
Qy 463 TAGAGACTGGGTGTGAAGTAGACACGACACAAACAGATAAGCAATGAAGAACAT 522
Db 636 TAAAGACTGGCGTTTGCACACTGGATACCGGACGCTGACTGACGCGATCCTCAGCAGGG 695
Qy 523 TGGCGGCGCGTTAGATACGCTGAGCCAGCGCTACGCTAAACCTTTTGCACAGATGGGG 582
Db 696 CAGGAGGCTCAATTGCTGACTTTACCGGGCATCGGCAACGGCGTTTCGCGAATCGAAC 755
Qy 583 ACCTGCTCAATTTTGGCGCTTCTCTTATTTGCAATCTTTGCAACAGACAGGAAACGCT 642
Db 756 GGGTGTCTTAATTTTCGCGCAATCAAACTTGTCCCTTAAACGCTGAGAAACAGACGAAAGCT 815
Qy 643 GTGATTTTGGCCACTTTTGGCGCCCAATGAAGTTAACTGTTAATAAGAGAGGACAAAAGTGA 702
Db 816 GTTCATTAACGAGGCAATACCATCGGAATCAAGGTGAGCGCGA-----CAATGCTC 869
Qy 703 CCCTCAGTGGCGCACTGGCGTTATCAGCAATTTGGGTGAAATCTTCTTATTAACAAACG 762
Db 870 CATTAACCGGTGGCGTAAAGCCTCGCATCAATGTCTGACGGAGATATTTCTCTGCAACAG 929
Qy 763 CACAAGCCATGCCAGAGGTTGCCCTGGCAACCGCTAAAGGGGGGAGAAATTTGGGTATCCT 822
Db 930 CACAGGNAATGCCGAGCGCGGTGGGAGAGATCACCGATTACACAGTGGAAACCT 989
Qy 823 TATTGTCAATACACGCGCAATTAATTTAATTTAATGCGAAACACCGTATATCCCGGTC 882
Db 990 TGCTAAGTTTGCATAACGCGCAATTTTATTTGCTACAAACGCAACCGCAGAGGTTGCCGCA 1049
Qy 883 ATAAAGGAGCGCCATTTATCAGCAGATAGTACGGCTTTAAACCTTCACTGATGCTC 942
Db 1050 GCCCGGCCACCCGTTATAGATTGATTAAGATTAAGACAGCGTTGACGCCCATC---CACCGC 1106
Qy 943 AGGGGCAAAAGCTACCCATTTACGCCCAAAACCGGCTTTGTTCTCGTGGCGCATGATA 1002
Db 1107 AAAAAGAGCGGTATGGTGTGACATTAACCACTTCACTGCTTTTATCGCGGACACGATA 1166
Qy 1003 CCATATGCGCAATATTGCGGTATGCTGGGAGCGGAGTGGAGTACCGGACCGCAACCTG 1062
Db 1167 CTAATCTGGCAAAATCTCGGCGGCGCACTGGAGTCAACTGAGCTTCCCGGTGAGCGCG 1226
Qy 1063 ATAATACTCCACAGTGGGGGATGGTTTGTGAACTATGCGCAAAATCCCGGATAACCAAC 1122
Db 1227 ATAACAGCGCGCAGGTGGTGAACCTGGTTTGAACGCTGGCGTGGCTAAGCGATAACA 1286
Qy 1123 AGCCTAGTTCGCGTGAAGATGTTCTACCAACGATGGATCAGTTGCGTAAT 1175
Db 1287 GCCAGTGGATTGAGTTTCGCTGCTCTCCAGACTTTACAGCAGATGCGTAT 1339

RESULT 14
AX042376
LOCUS AX042376 3470 bp DNA linear PAT 23-NOV-2000
DEFINITION Sequence 5 from Patent WO0064247.
ACCESSION AX042376
VERSION AX042376.1 GI:11340994
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
1 artificial sequences.
REFERENCE
1 Forsberg,C.W., Golovan,S. and Phillips,J.P.
AUTHORS Transgenic animals expressing salivary proteins
TITLE Patent: WO 0064247-A 5 02-NOV-2000;
JOURNAL UNIVERSITY OF GUELPH (CA)
FEATURES
Location/Qualifiers
1..3470
/organism="synthetic construct"

/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="R15/APPA transgene"

Query Match 17.0%; Score 224.6; DB 6; Length 3470;
Best Local Similarity 53.2%; Pred. No. 1.2e-55;
Matches 571; Conservative 0; Mismatches 489; Indels 13; Gaps 4;

ORIGIN

Qy 107 TAGAAGCTGTGGTTATTTTGTAGTCCGCATCGTGTTCGTCGCGACCAA-ACAAGCGACG 165
Db 1899 TGGAAAGTGTGGTATTTGTCTGTCGTGTCGTGTCCTCAACCAAGCGCACGCAAC 1958
Qy 166 TTATGATGATGTTAGCCGACAGATAGTGGCCCAATAGCCCGTAAAGCGGGGTATTTAA 225
Db 1959 TGATGAGGATGTCAACCCAGACGATAGCCCAACCTGCGCGGTAAACCTGGGTGGCTGA 2018
Qy 226 CCGCACGTCGTCGAGTTGGTCAATGATCGGGGGTTTTATGCTGATTACTTTTCGCA 285
Db 2019 CACCGCGGTGGTGGCTAATCGCTATCTCGGACATTTACCAACGCGAGCTCTGGTAG 2078
Qy 286 GCGTGTGTTTGTAGCG---GCGGATGTCCGCGAGAGGGGGGGGTATATGACAGGCGAG 342
Db 2079 CCGACGGAATTTGCGGAAAGGGCTGCCCGCAGTCTGCTCAGTCCGCAATTTATGCTG 2138
Qy 343 ATATCGATCAACGTCACCGCTTAAACCGGACAGGCAATTTCTTGATGTTGGCTCCGGGT 402
Db 2139 ATGTGACGAGGCTACCGGTAAACAGGGGAGCCCTTGCCTCCCGGCTGGCACCTGACT 2198
Qy 403 GTGTTTGAACCTGTGATATCAGGCGGATTTGAAAGAACCGATCCCTGTTCATCCGG 462
Db 2199 GTGCAATAACCTGTACATACCCAGGCGAGATAGTCCAGTCCCGATCCGTTATTTAATCCTC 2258
Qy 463 TAGAGACTGGCGTGTGAAGTTAGACAGCAGCAACAGATAAACCGATTGAAGAACGAT 522
Db 2259 TAAAGACTGGCTTTGCCAATCTGGATTAACGGAACGTAAGTGAACCGATCTCTCAGAGG 2318
Qy 523 TGGCGGCGCGTTAGATACGCTGAGCGCTACGCTAAACCTTTTCCCGAGATGGGG 582
Db 2319 CAGGAGGGTCAATTTGCTGACTTTACGGGCGATCGGCAACCGCGTTCGCGAATCGAAC 2378
Qy 583 AGTCTGATTTTGGCGCTTCTCTTATTCGAATCTTTGCAACAGCAAGGAAACGCT 642
Db 2379 GGGTCTTAATTTTCCGCAATCAAACTTGTGCTTTAAACGCTGAGAAACAGGACGAAAGCT 2438
Qy 643 GTGATTTTGCACCTTTGCGGCAATGAAGTTAACTGTTAATAAGAGAGGACAAAGTGA 702
Db 2439 GTTCATTAACGAGGCAATTAACATCGGAATCAAGGTGAGCGCGA-----CAATGCTC 2492
Qy 703 CCTCAGTGGCGCACTGGGTTATCATCGACATTTGGGTGAATCTTCTTATTAACAAACG 762
Db 2493 CATTAACCGGTGCGGTAAAGCCTCGCATCAATGCTGACGAGATATTTCTCTGCAACAG 2552
Qy 763 CACAAGCCATGCCAGAGTTGCTCGCAACCGCTAAAGGGGCGGAGAAATTTGGGTATCCT 822
Db 2553 CACAGGAATGCCGAGCGGGTGGGAGGATCACCGATTCAACAGTGGACACCT 2612
Qy 823 TATTGTCAATTAATAACCGCAATTTAATTTAATGGCAAAACACCGTATATTCGCCCGTC 882
Db 2613 TGCTAAGTTTGCATAACCGCAATTTTATTTGCTACAAACGCGACGCGAGAGTTGCCGCA 2672
Qy 883 ATAAAGGAGCGCATTTATTAACAGCAGATAGATAGCGCTTTAACCTTCAACTGGATGCTC 942
Db 2673 GCGGCGCAACCCGTTTATAGTTTGAATCAAGACAGCGTTGACGCCCATC---CACGCG 2729
Qy 943 AGGGGCAAAAGCTACCCATTTTACGCCCAAAACCGGCTCTTGTTCCTCGGTGGGCAATGATA 1002
Db 2730 AAAAAGAGCGGTATGGTGTGACATTAACCACTTCACTGCTGTTTATCGCGGACACGATA 2789
Qy 1003 CCAATATTGCCAATATTGCGGTATGCTGGGAGCGGAGTGGCAGCTACCGGACGACCTG 1062
Db 2790 CTAACTGTGCAAAATCTCGCGCGCGCACTGGAGCTCAACTGGACGCTTCCCGGTACGCCG 2849

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2004, 19:58:08 ; Search time 378.238 seconds

(without alignments)
14881.793 Million cell updates/sec

Title: US-10-021-723B-3

Perfect score: 1325

Sequence: 1 atgtcagattagaaatcg.....aacctgctgccacatctaa 1325

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 337363 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_29Jan04.*
1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001as.*
5: Geneseq2001bs.*
6: Geneseq2002s.*
7: Geneseq2003as.*
8: Geneseq2003bs.*
9: Geneseq2003cs.*
10: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1314	99.2	1326	6	ABQ73843 Versinia
2	1306	98.6	1326	7	ACA53738 Prokaryot
3	1304.2	98.4	1326	6	ABQ73842 Versinia
4	827.6	62.5	1326	6	ABQ73847 Versinia
5	242	18.3	1281	8	ACF35789 Nov9X phy
6	226.4	17.1	5421	3	ACA68299 SV40/APPA
7	226.2	17.1	1308	9	ADC87742 DNA encod
8	225.2	17.0	1901	6	AAD25463 Escherich
9	225.2	17.0	1901	7	ADA19452 E. coli K
10	225.2	17.0	1901	7	ADA19449 E. coli K
11	225.2	17.0	1901	9	ADC87744 DNA encod
12	224.6	17.0	1299	3	ABK12514 DNA encod
13	224.6	17.0	1299	7	ACA19297 Prokaryot
14	224.6	17.0	3470	3	ACA68298 R15/APPA
15	224.6	17.0	4060	3	ACA68296 R15/APPA
16	224.6	17.0	6116	3	ACA68297 R15/APPA
17	224.6	17.0	6708	3	ACA68295 R15/APPA
18	224.6	17.0	17732	3	ACA68300 Lama2/APP
19	224.6	17.0	20623	3	ACA68294 Lama2/APP
20	221.4	16.7	1486	4	AAD06831 E. coli a
21	218.8	16.5	1323	5	ACA88985 Escherich
22	218.8	16.5	1323	6	AAD36473 Escherich
23	218.8	16.5	1323	6	AAD25460 Escherich

24 218.8 16.5 1323 7 ADA19445
25 218.2 16.5 1486 4 AAD06832 E. coli a
26 218.2 16.5 1486 7 ACC57873 Escherich
27 218.2 16.5 1489 3 AAA28216 E. coli a
28 218.2 16.5 1489 7 ACC57872 Escherich
29 177.8 13.4 1256 8 AAD57147 Maize-opt
30 177.8 13.4 1256 8 AAD57148 pNOV4054
31 177.4 13.4 1331 8 AAD57149 Versinia
32 177.4 13.4 1431 6 ABQ73846 Versinia
33 143.6 10.8 1272 2 AAX26540 DNA encod
34 133.2 10.1 846 4 AAS41940 Genomic s
35 104.4 7.9 1737 5 AAS89885 DNA encod
36 100.4 7.6 519 4 AAS41139 DNA encod
37 73.2 5.5 466 7 ACA15311 Prokaryot
38 68.4 5.2 11710 6 AAH43718 E. coli g
39 65.8 5.0 1264 5 AAS88443 DNA encod
40 65.8 5.0 1264 5 AAS94316 DNA encod
41 64.6 4.9 1230 6 ABQ73845 Rhizobium
42 59.8 4.5 1266 6 ABQ73844 Versinia
43 50.2 3.8 2000 7 ADA71938 Rice Gene
44 48.2 3.6 4050 5 AAS64582 DNA encod
45 48 3.6 5975 2 AAQ55758 Klebsiell

ALIGNMENTS

RESULT 1

ABQ73843
ID ABQ73843 standard; DNA; 1326 BP.
XX
AC ABQ73843;
XX
DT 08-OCT-2002 (first entry)
XX
DE Versinia pestis phytase nucleotide sequence SEQ ID NO:3.
XX
KW Versinia pestis; phytase; enzyme; E.C.3.1.3.8; phytate; foodstuff; feed;
KW gene; ds.
XX
OS Versinia pestis.
XX
FH Key Location/Qualifiers
FT CDS 1..1260
FT /tag= a
FT /partial
FT /EC_number= "3.1.3.8"
FT /product= "phytase"
FT /note= "no stop codon given"

WO200248332-A2.

20-JUN-2002.

12-DEC-2001; 2001WO-US048774.

12-DEC-2000; 2000US-0255090P.

(DIVE-) DIVERSA CORP.

Short J, Mathur EJ, Richardson T, Robertson D, Barton N;

WPI; 2002-583504/62.

P-PSDB; ABP51930.

Novel recombinant phytase protein and polynucleotide for improving nutritional value of phytate-containing foodstuff, in animal feed and feed supplements and to degrade excess phytase from environment or sample.

Claim 1; Fig 5C; 208pp; English.

The present invention describes an isolated phytase protein (I). (I) can

CC be used for improving the nutritional value of a phytate-containing
CC foodstuff, by contacting the phytate-containing foodstuff with (i), where
CC the enzyme catalyses the liberation of inorganic phosphate from the
CC phytate-containing foodstuff, and so improving the nutritive value of the
CC contacted foodstuff. The liberation of inorganic phosphate occurs prior
CC to or after the ingestion of phytate-containing foodstuff by a recipient
CC organism. Nucleotide sequences (ii) encoding (i) can be used for
CC producing an animal feed, by transforming a plant, plant portion or plant
CC cell with a nucleic acid expression vector, comprising (ii), culturing
CC the plant, plant portion or plant cell under conditions in which the
CC phytase protein is expressed and converting the plant, plant portion or
CC plant cell into a composition suitable for animal feed. The animal is
CC preferably a monogastric animal or a ruminant. (i) and (ii) are useful in
CC animal feed and feed supplements as well as in treatments to degrade or
CC remove excess phytate from the environment or a sample. (i) reduces
CC phytate levels in animal manure and so reduces the phosphate pollution of
CC the environment. The present sequence encodes a phytase from the present
CC invention
XX
SQ Sequence 1326 BP; 351 A; 285 C; 359 G; 331 T; 0 U; 0 Other;

Query Match 99.2%; Score 1314; DB 6; Length 1326;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1325; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 1 ATGTCAGTATTAGAAATCGTGACGGCTATCTGGCTGGTATTGATGCTAAGCGGATTG 60
DB 1 ATGTCAGTATTAGAAATCGTGACGGCTATCTGGCTGGTATTGATGCTAAGCGGATTG 60
QY 61 GCTGCTATTACTCGCGCGTAGCCGCGAGCCATCGGGCTATCTTAAAGCGTGTGTT 120
DB 61 GCTGCTATTACTCGCGCGTAGCCGCGAGCCATCGGGCTATCTTAAAGCGTGTGTT 120
QY 121 ATTTGAGTCGCCATGCTGTTCTCGCCGACCAAC-AACGAGCGTTATGATGATGTT 179
DB 121 ATTTGAGTCGCCATGCTGTTCTCGCCGACCAACAAACGAGCTTATGATGATGTT 180
QY 180 ACSCCAGATAAGTGCCACATGCGCGGTAAAGCGGGTATTAAAGCCACGTTGGTGGC 239
DB 181 ACSCCAGATAAGTGCCACATGCGCGGTAAAGCGGGTATTAAAGCCACGTTGGTGGC 240
QY 240 GAGTGTGTCATTTGAGTGGGGGGTATTATGGTGATTCTTCGAGCGCTTGGTTGTTA 299
DB 241 GAGTGTGTCATTTGAGTGGGGGGTATTATGGTGATTCTTCGAGCGCTTGGTTGTTA 300
QY 300 GCGCGGGGATGTCGCGCAGAGGGGGGGTATTATGCAAGCAGATATCGATCAACGTACC 359
DB 301 GCGCGGGGATGTCGCGCAGAGGGGGGGTATTATGCAAGCAGATATCGATCAACGTACC 360
QY 360 CGCTTAACCGGACAGGCAATTTCTTGATGGTGCGTCCGGGGTGGTGGTACCGTGCAT 419
DB 361 CGCTTAACCGGACAGGCAATTTCTTGATGGTGCGTCCGGGGTGGTGGTACCGTGCAT 420
QY 420 AATCAGCCGATTGAAAAAGACCGATCCCTGTTCCATCCGCTAGAGACTGGCGTGTGT 479
DB 421 AATCAGCCGATTGAAAAAGACCGATCCCTGTTCCATCCGCTAGAGACTGGCGTGTGT 480
QY 480 AAGTTAGACACGACACACAGATAAGCGATTGAAAGCAGATTGGCGCGCGCTTAGAT 539
DB 481 AAGTTAGACACGACACACAGATAAGCGATTGAAAGCAGATTGGCGCGCGCTTAGAT 540
QY 540 ACGGTGAGCCAGCGCTACGCTAAACCTTTTCCAGATGGGGGACGCTGCTGAATTTGCG 599
DB 541 ACGGTGAGCCAGCGCTACGCTAAACCTTTTCCAGATGGGGGACGCTGCTGAATTTGCG 600
QY 600 GCTTCTCTTATTGCAATCTTTGCAACGACGAGGAAACGCTGATTTTGGCCACTTT 659
DB 601 GCTTCTCTTATTGCAATCTTTGCAACGACGAGGAAACGCTGATTTTGGCCACTTT 660
QY 660 GCGGCCAATGAAGTTAAGCTTAATAAGAGAGGACAAAGTGAACCTTCAGTGGGCCACTG 719
DB 661 GCGGCCAATGAAGTTAAGCTTAATAAGAGAGGACAAAGTGAACCTTCAGTGGGCCACTG 720

QY 720 GCGTTATATCGACATTGGGTGAATCTTCTTATTATCAAAACGACAGCCATGCCAGAG 779
DB 721 GCGTTATATCGACATTGGGTGAATCTTCTTATTATCAAAACGACAGCCATGCCAGAG 780
QY 780 GTTCCCTGGCAACGGCTAAAAAGGGGGGAGAAATGGGTATCCTTATTGTTCATTACATAAC 839
DB 781 GTTCCCTGGCAACGGCTAAAAAGGGGGGAGAAATGGGTATCCTTATTGTTCATTACATAAC 840
QY 840 GCGCAATTTAATTTAATGGCAAAACACCGGTATATGCGCGCTATTAAGGGAGCGCATTA 899
DB 841 GCGCAATTTAATTTAATGGCAAAACACCGGTATATGCGCGCTATTAAGGGAGCGCATTA 900
QY 900 TTACAGCAGATAGATAGCGCTTTTAAACCCCTTCAACTGGATGCTCAGGGGCAAAAGCTACCC 959
DB 901 TTACAGCAGATAGATAGCGCTTTTAAACCCCTTCAACTGGATGCTCAGGGGCAAAAGCTACCC 960
QY 960 ATTTACGCCCAAAACCGGGTCTTGTTCCTCGGTGGGATGATACCAATATTGCAATATT 1019
DB 961 ATTTACGCCCAAAACCGGGTCTTGTTCCTCGGTGGGATGATACCAATATTGCAATATT 1020
QY 1020 GCGGGTATGCTGGGAGCGGACTGGCAGCTACCCGAGCAACCTGATATATCTCCACCAGGT 1079
DB 1021 GCGGGTATGCTGGGAGCGGACTGGCAGCTACCCGAGCAACCTGATATATCTCCACCAGGT 1080
QY 1080 GGGGGATTGGTTTTTGAACCTATGCGCAAAATCCCGATAACCCAGCGCTACGTTGCGGTG 1139
DB 1081 GGGGGATTGGTTTTTGAACCTATGCGCAAAATCCCGATAACCCAGCGCTACGTTGCGGTG 1140
QY 1140 AAGATGTTCTACCAACGATGGATCAGTTCGCTAATGCGCAAAATTTGGATCTGAAAAAT 1199
DB 1141 AAGATGTTCTACCAACGATGGATCAGTTCGCTAATGCGCAAAATTTGGATCTGAAAAAT 1200
QY 1200 AACCCAGCGGTATTATTTCGTTGCGAGTTGCTGGTTGCGAAATAAACGTCGACGATAAG 1259
DB 1201 AACCCAGCGGTATTATTTCGTTGCGAGTTGCTGGTTGCGAAATAAACGTCGACGATAAG 1260
QY 1260 CTTTTCGAGCTTGATATATTCCCAAGAGAGTGGCTAAGGTAAATTGAACCTGCTGCCAC 1319
DB 1261 CTTTTCGAGCTTGATATATTCCCAAGAGAGTGGCTAAGGTAAATTGAACCTGCTGCCAC 1320
QY 1320 ATCTAA 1325
DB 1321 ATCTAA 1326

RESULT 2
ACA53738
ID ACA53738 standard; DNA; 1326 BP.
XX ACA53738;
AC ACA53738;
DT 19-JUN-2003 (first entry).
XX Prokaryotic essential gene #35395.
DE Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX Versinia pestis.
XX WO200277183-A2.
FN 03-OCT-2002.
XX 21-MAR-2002; 2002WO-US009107.
XX 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX WPI; 2003-029926/02.
 DR P-PSDB; ABU49868.
 XX
 XX New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 XX Claim 14; SEQ ID NO 41608; 1766pp; English.
 XX
 CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
 CC prokaryotic essential genes. Note: the sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 XX ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 1326 BP; 348 A; 286 C; 361 G; 331 T; 0 U; 0 Other;

Query Match 98.6%; Score 1306; DB 7; Length 1326;
 Best Local Similarity 99.5%; Pred. No. 0;
 Matches 1320; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 1 ATGTCAGTATTAGAAATCGGTACGGCTATCTGGCTGGTATTGATGCTAAGCGGATTG 60
 DB 1 ATGTCAGTATTAGAAATCGGTACGGCTATCTGGCTGGTATTGATGCTAAGCGGATTG 60

QY 61 GCTGCTATTACTGCGCGGTTAGCGCGAGCCATCGGGCTATACCTTTAGAACGTTGGTT 120
 DB 61 GCTGCTATTACTGCGCGGTTAGCGCGAGCCATCGGGCTATACCTTTAGAACGTTGGTT 120

QY 121 ATTTTGGTCCCATGTTGTTGCTCCCGACCAAC- AACGACGCTTATGATGATGTT 179
 DB 121 ATTTTGGTCCCATGTTGTTGCTCCCGACCAACAAACGACGCTTATGATGATGTT 180

QY 180 ACGCCAGATAAGTGCCCAATAGCGCGGTAAAGCGGGTATTTAACGCCACGTTGGTGG 239
 DB 181 ACGCCAGATAAGTGCCCAATAGCGCGGTAAAGCGGGTATTTAACGCCACGTTGGTGG 240

QY 240 GAGTTGGTCATGATGGGGGGGTTTATGGTACTTTCGACGCTTGGTTGTTA 299
 DB 241 GAGTTGGTCATGATGGGGGGGTTTATGGTACTTTCGACGCTTGGTTGTTA 300

QY 300 CGCGCGGATGTCGCGCAGAGGGGGGTATATGACAGGCGAGATATCGATCAACGTACC 359

DB 301 GCGCGGGATGTCGCGCAGAGGGGGGTATATGACAGGCGAGATATCGATCAACGTACC 360
 QY 360 CGCTTAACCGGACAGGCATTTCTTATGATGGTGGGTCGCGGGTGGTGGTGGAT 419
 DB 361 CGCTTAACCGGACAGGCATTTCTTATGATGGTGGGTCGCGGGTGGTGGTGGAT 420
 QY 420 AATCAGCGCGATTTGAAAAGACCGATCCCTGTTCCATCCGTTAGAGACGTCGCGTGT 479
 DB 421 AATCAGCGCGATTTGAAAAGACCGATCCCTGTTCCATCCGTTAGAGACGTCGCGTGT 480
 QY 480 AAGTTAGACAACGCCAACAACAGATTAAGGATTAAGACGATTTGGGCGGGCGGTAGAT 539
 DB 481 AAGTTAGACGCCGCAACACAGATTAAGGATTAAGACGATTTGGGCGGGCGGTAGAT 540
 QY 540 ACGGTGAGCGCAGCGGTACGCTAAACCTTTTGGCCAGATGGGCGAGCTGTGTAATTTGG 599
 DB 541 ACGGTGAGCGCAGCGGTACGCTAAACCTTTTGGCCAGATGGGCGAGCTGTGTAATTTGG 600
 QY 600 GCTTCTCTTATGCAATCTTTGCAACAGCAAGAAAACGTTGATTTTCCCACTTT 659
 DB 601 GCTTCTCTTATGCAATCTTTGCAACAGCAAGAAAACGTTGATTTTCCCACTTT 660
 QY 660 GGGGCCAATGAAGTTAACTTAATAAGAGGACAAAAGTCACTCAGTGGGCGCACTG 719
 DB 661 GGGGCCAATGAAGTTAACTTAATAAGAGGACAAAAGTCACTCAGTGGGCGCACTG 720
 QY 720 GGGTTATCATCGACATGCGGTGAATCTTTTATTAACAAACGCAAGCCATGCGAGAG 779
 DB 721 GGGTTATCATCGACATGCGGTGAATCTTTTATTAACAAACGCAAGCCATGCGAGAG 780
 QY 780 GTTGCTGCGCAACGGCTAAAGGGCGGAGAAATTCGGTATCTCTTATTTGTCATTACATAC 839
 DB 781 GTTGCTGCGCAACGGCTAAAGGGCGGAGAAATTCGGTATCTCTTATTTGTCATTACATAC 840
 QY 840 GCGCAATTTAATTAATGCAAAAAACCGTATATGCGCGGTCATAAAGGGACGCGCATTA 900
 DB 841 GCGCAATTTAATTAATGCAAAAAACCGTATATGCGCGGTCATAAAGGGACGCGCATTA 900
 QY 900 TTACAGCAGATAGTACGGCTTTAACCTTCACTGGATGCTCAGGGGCAAAAGCTACCC 959
 DB 901 TTACAGCAGATAGTACGGCTTTAACCTTCACTGGATGCTCAGGGGCAAAAGCTACCC 960
 QY 960 ATTTTCAGCGCCAAAACCGGGTCTTTGTTCTCGTGGGCGATGATACCAATATTGCGCAATAT 1019
 DB 961 ATTTTCAGCGCCAAAACCGGGTCTTTGTTCTCGTGGGCGATGATACCAATATTGCGCAATAT 1020
 QY 1020 GCGGTATGCTGGAGCGGACCTGGGCGAGCTACCGGACCACTGATTAATCTCCACAGGT 1079
 DB 1021 GCGGTATGCTGGAGCGGACCTGGGCGAGCTACCGGACCACTGATTAATCTCCACAGGT 1080
 QY 1080 GGGGGATTGGTTTTTGAACCTATGCGCAAAATCCGGATAACCAACGCGCTACGTTGCGGTG 1139
 DB 1081 GGGGGATTGGTTTTTGAACCTATGCGCAAAATCCGGATAACCAACGCGCTACGTTGCGGTG 1140
 QY 1140 AAGATGTTTACCAACGATGATCAGTTGGTATGCGGAAAAAATTTGATCTGAAAAAT 1199
 DB 1141 AAGATGTTTACCAACGATGATCAGTTGGTATGCGGAAAAAATTTGATCTGAAAAAT 1200
 QY 1200 AACCAGCGGGTATTATTTCGTTGCTGCTGTTGTAATAAATGCGTACCGATAAG 1259
 DB 1201 AACCAGCGGGTATTATTTCGTTGCTGCTGTTGTAATAAATGCGTACCGATAAG 1260
 QY 1260 CTTTGGAGCTTGATACATTTCAAAAGAAAGTGGCTAAGTAAATTTGAACCTGCTGCCAC 1319
 DB 1261 CTTTGGAGCTTGATACATTTCAAAAGAAAGTGGCTAAGTAAATTTGAACCTGCTGCCAC 1320
 QY 1320 ATCTAA 1325
 DB 1321 ATCTAA 1326

ID ABQ73842 standard; DNA; 1326 BP.
AC ABQ73842;
XX
XX
DT 08-OCT-2002 (first entry)
XX
DE Yersinia pestis phytase nucleotide sequence SEQ ID NO:1.
XX
XX Yersinia pestis; phytase; enzyme; E.C.3.1.3.8; phytate; foodstuff; feed;
KW gene; ds.
XX
XX Yersinia pestis.
XX
XX
FH Key Location/Qualifiers
FT CDS 1..1326
FT FT /*tag= a
FT FT /EC number= "3.1.3.8"
FT FT /product= "phytase"
FT FT /trans_except= (pos:469..471,aa:Xaa)
FT FT /trans_except= (pos:487..489,aa:Xaa)
FT FT /trans_except= (pos:490..492,aa:Xaa)
FT FT /trans_except= (pos:520..522,aa:Xaa)
FT FT /note= "Xaa is unspecified"
XX
XX WO200248332-A2.
XX
XX
XX
PD 20-JUN-2002.
XX
XX
PF 12-DEC-2001; 2001WO-US048774.
XX
XX
PR 12-DEC-2000; 2000US-0255090P.
XX
XX (DIVE-) DIVERSA CORP.
XX
XX
XX Short J, Mathur EJ, Richardson T, Robertson D, Barton N;
XX
XX WPI; 2002-583504/62.
DR P-PSDB; ABP51929.
XX
XX
XX Novel recombinant phytase protein and polynucleotide for improving
PT nutritional value of phytate-containing foodstuff, in animal feed and
PT feed supplements and to degrade excess phytase from environment or
PT sample.
XX
XX
XX Claim 1; Fig 5A; 208pp; English.
XX
XX
XX The present invention describes an isolated phytase protein (I). (I) can
CC be used for improving the nutritional value of a phytate-containing
CC foodstuff, by contacting the phytate-containing foodstuff with (I), where
CC the enzyme catalyses the liberation of inorganic phosphate from the
CC phytate-containing foodstuff, and so improving the nutritive value of the
CC contacted foodstuff. The liberation of inorganic phosphate occurs prior
CC to or after the ingestion of phytate-containing foodstuff by a recipient
CC organism. Nucleotide sequences (II) encoding (I) can be used for
CC producing an animal feed, by transforming a plant, plant portion or plant
CC cell with a nucleic acid expression vector, comprising (II), culturing
CC the plant, plant portion or plant cell under conditions in which the
CC phytase protein is expressed and converting the plant, plant portion or
CC plant cell into a composition suitable for animal feed. The animal is
CC preferably a monogastric animal or a ruminant. (I) and (II) are useful in
CC animal feed and feed supplements as well as in treatments to degrade or
CC remove excess phytate from the environment or a sample. (I) reduces
CC phytate levels in animal manure and so reduces the phosphate pollution of
CC the environment. The present sequence encodes a phytase from the present
XX invention
SQ Sequence 1326 BP; 348 A; 283 C; 359 G; 331 T; 0 U; 5 Other;
Query Match 98.4%; Score 1304.2; DB 6; Length 1326;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 1317; Conservative 0; Mismatches 8; Indels 1; Gaps 1;
XX 1 ATGTCAGTATTAGAAAATCGTGTACGGCTATCTCGGCTGTGATGTGCTAACGGGATTG 60

1 ATGTCAGTATTAGAAAATCGTGTACGGCTATCTGGGCTGTGATGTGCTAACGGGATTG 60
61 GCTGCTATTACTGCGCGGTAGCCGCGGAGCATCGGGCTATCTTTAGAACGTGTGTT 120
61 GCTGCTATTACTGCGCGGTAGCCGCGGAGCATCGGGCTATCTTTAGAACGTGTGTT 120
121 ATTTTGTAGTCCCATGTGTTTGTGCTCCGCCAACCAAC- AACGCAGCTTATGAATGATGTT 179
121 ATTTTGTAGTCCCATGTGTTTGTGCTCCGCCAACCAACAAACGACGCTTATGAATGATGTT 180
180 ACGCCAGATAAGTGGCCCAATGCGCGGTAAAGCGGGGTATTTAAACGCCACGTGGTGGC 239
181 ACGCCAGATAAGTGGCCCAATGCGCGGTAAAGCGGGGTATTTAAACGCCACGTGGTGGC 240
240 GAGTTGTGCATATTGATGGGGGGGTTTATGTTGATTAATCTTTCGACGCCCTTGTGTTTA 299
241 GAGTTGTGCATATTGATGGGGGGGTTTATGTTGATTAATCTTTCGACGCCCTTGTGTTTA 300
300 GCGCGGGATGTCGCGCAGAGGGGGGGGTATATGACAGGAGATATGATCAACGTAAC 359
301 GCGCGGGATGTCGCGCAGAGGGGGGGGTATATGACAGGAGATATGATCAACGTAAC 360
360 CGTTAACCGGACAGGCAATTTCTTGTGTTGCTCGGGGTGTTGTTGACCGTGCAT 419
361 CGTTAACCGGACAGGCAATTTCTTGTGTTGCTCGGGGTGTTGTTGACCGTGCAT 420
420 AATCAGGCCGATTTGAAAAGACCGGATCCCTGTTCCATCCGCTAGAGAGCTGGCGTGTGT 479
421 AATCAGGCCGATTTGAAAAGACCGGATCCCTGTTCCATCCGCTAGAGAGTGGCGTGTGT 480
480 AAGTTAGAACGACGACAAACAGATAAGAGCATTAAGAAAGCATTTGGCGGCGCGTATAGAT 539
481 AAGTTANACGNGACAAACAGATAAGAGCATTAAGAAAGCATTTGGCGGCGCGTATAGAT 540
540 ACGGTGAGCCAGCGCTACGCTAAACCTTTTGCAGATGGGGACGCTGTAATTTTGGC 599
541 ACGGTGAGCCAGCGCTACGCTAAACCTTTTGCAGATGGGGACGCTGTAATTTTGGC 600
600 GCTTCTCCTTATTCGCAATCTTTGCAACAGCAAGGAAACGTTGATTTTGGCCACTTT 659
601 GCTTCTCCTTATTCGCAATCTTTGCAACAGCAAGGAAACGTTGATTTTGGCCACTTT 660
660 GCGGCCAATGAAGTTAAAGTTTAAAGAAAGGGAACAAAGTGACCCCTCAGTGGGCCACTG 719
661 GCGGCCAATGAAGTTAAAGTTTAAAGAAAGGGAACAAAGTGACCCCTCAGTGGGCCACTG 720
720 GCGTTATCATCGACATTTGGGTGAATCTTCTTATTAACAAACGACAGCCATGCCAGAG 779
721 GCGTTATCATCGACATTTGGGTGAATCTTCTTATTAACAAACGACAGCCATGCCAGAG 780
780 GTTGCTCTGGCAACCGCTAAAGAGGGGGCGGAGAAATTTGGGTATCTTATTTGTCATTACATAAC 839
781 GTTGCTCTGGCAACCGCTAAAGAGGGGGCGGAGAAATTTGGGTATCTTATTTGTCATTACATAAC 840
840 GCGCAATTTAATTTAATGGGCAAAACACCGTATATCCCGTCAATAAGGGAGCGCCATTA 899
841 GCGCAATTTAATTTAATGGGCAAAACACCGTATATCCCGTCAATAAGGGAGCGCCATTA 900
900 TTACAGCAGATAGATAGCGCTTTTAAACCTTCAACTGGATGCTCAGGCGGCAAAAGCTACCC 959
901 TTACAGCAGATAGATAGCGCTTTTAAACCTTCAACTGGATGCTCAGGCGGCAAAAGCTACCC 960
960 ATTTGAGCCCAAAACCGGGTCTTGTTCCTCGGTGGGATGATACCAATATTTGCCAATATT 1019
961 ATTTGAGCCCAAAACCGGGTCTTGTTCCTCGGTGGGATGATACCAATATTTGCCAATATT 1020
1020 GCGGGTATGCTGGGAGCGGACTGGCAGCTACCCGAGCAACCTGATATACTTCCACCCAGGT 1079
1021 GCGGGTATGCTGGGAGCGGACTGGCAGCTACCCGAGCAACCTGATATACTTCCACCCAGGT 1080
1080 GCGGGATTTGGTTTTTGAACACTATGGCAAAATCCCGATTAACCAACGCGCTAGCTTGGCGTG 1139

Db 841 GCACAGTTCACCTGATGCTAAACGCCATACATTCGACGCCACAAAGGACGCGGCTT 900
Qy 900 TTACACGAGTACAGCTTTAAACCTTCACTGATGCTCAGGGCACAAGCTACCC 959
Db 901 TTACAGCAATCATACCGCACTGACCTTGCACCTGACCCAGGGCAAAATCTGCCG 960
Qy 960 ATTTCAGCCCAAAACCGGGTCTTTGTTCTCGGTGGCATGATACCAATATTCCTCAATATT 1019
Db 961 ATCTCGGCTCAGAACCGTGTATTTATCTCTGGGTGGCCACGACACAAATATTGCTAACATC 1020
Qy 1020 GCGGTATGCTGGAGCCGACCTGGAGCTTACCGAGCAACCTGATTAATCTCCACGAGT 1079
Db 1021 GCGGTATGCTGGGCGCAGATTGGCAGTTACCGGAACAACCGGATTAACACCCCGGGC 1080
Qy 1080 GGGGATTTGTTTGAACATATGCGCAAAATCCGGAATAACACAGCGCTACGTTGGGGTG 1139
Db 1081 GCGGTCTGCTTTGAGCTGTGGCAGATCCGGAATCCGGAATCATCAAGTTATGTGGCGTT 1140
Qy 1140 AGATGTTCTACAAAGATGATCAGTTCGCTGATGCGGAATTCGGAATTCGATCGAAAT 1199
Db 1141 AGATGTTCTACAGACCATGGATCAATTGCGTAACCGCGAGAGCTGGATTTAAGAAC 1200
Qy 1200 AACCCAGCGGTATTTTCCGTTGCGATGCTGCTGTTGTGAAATTAACCGTGACGATAAG 1259
Db 1201 AATCCCGCGCATCATCATGTCGCTGTGGCGGCTGCGAGATTAATGTCGACGATAA 1260
Qy 1260 CTTTGGAGCTTGATCATTCGCAAAAGAGAGTGGTAAAGTAACTGAACTGCTGCCAC 1319
Db 1261 CTGTGGCACTTGATATCTTTTCAAAAAGAGTAGCGAAATCATTTGAACCTGCTGTCAT 1320
Qy 1320 ATCTAA 1325
Db 1321 ATTAA 1326

RESULT 5
ACF35789
ID ACF35789 standard; DNA; 1281 BP.
AC ACF35789;
XX ACF35789;
DT 06-NOV-2003 (first entry)
XX Nov9X phytase polypeptide encoding DNA.
DE Nov9X; thermotolerant; phytase; phosphate; animal food; phosphorous;
KW gene; ds.
XX Synthetic.
OS
FH Key Location/Qualifiers
CDS 25..1263
FT /*tag= a
FT /transl_except= (pos; 25..27, aa: Met)
FT /product= "Nov9X"
XX WO2003057247-A1.
FN 17-JUL-2003.
XX 30-DEC-2002; 2002WO-US041785.
PF 28-DEC-2001; 2001US-0344523P.
XX (SYGN) SYNGENTA PARTICIPATIONS AG.
FA (DIVE-) DIVERSA CORP.
XX Lanahan ML, Koepf E, Kretz K;
XX WPI; 2003-598328/56.
DR P-PSDB; ABR82310.
XX Preparing thermotolerant phytase useful for preparing animal feed or

human food, by expressing in microbial host cell expression cassette comprising a promoter operably linked to a nucleic acid encoding the enzyme.
XX Example 3; Page 77-78; 79pp; English.
XX The invention relates to preparing a thermotolerant phytase (I), involving expressing in a microbial host cell an expression cassette comprising a promoter operably linked to a nucleic acid molecule encoding (I). The thermotolerant phytase has a specific activity of greater than 400 U/mg, 600 U/mg or 800 U/mg at pH 4.5 and 37plusC. The nucleic acid molecule encodes a fusion polypeptide comprising the thermotolerant phytase and a signal sequence which is operably linked to the thermotolerant phytase. The thermotolerant phytase is glycosylated and has a simulated gastric half-life of greater than 25 minutes at a pH greater than 2.0 and less than 4.0. (I) is useful for preparing animal feed, improving nutritive value of animal feed and human food, and for preparing human food. A heat-treated animal feed mixture comprising an inorganic phosphate at below 0.45% and (I) is useful for reducing feed conversion ratios or increasing weight gain of animals fed diets with inorganic phosphate at levels below 0.45%. It is also useful for decreasing phosphate levels in excreta from an animal. An animal feed composition comprising (I) is useful for decreasing the feed conversion ratio and increasing the weight gain of an animal. The composition is useful for minimizing inorganic dietary requirements of phosphorous in an animal, and for enhancing organic phosphate utilization from organic phosphorous sources in feed for an animal, where the feed increases the bioavailability of inorganic phosphorous in the feed to the animal. It is also useful for decreasing phosphate levels in excreta from an animal. In all the above mentioned conditions, the feed is a poultry or swine feed. The thermotolerant phytase has a half-life of about 30 minutes in the digestive tract of the animal. The present sequence represents a thermotolerant Nov9X phytase encoding DNA
XX Sequence 1281 BP; 322 A; 295 C; 300 G; 364 T; 0 U; 0 Other;
Query Match 18.3%; Score 242; DB 8; Length 1281;
Best Local Similarity 52.5%; Pred. No. 6.7e-68;
Matches 628; Conservative 0; Mismatches 555; Indels 13; Gaps 4;

Qy 93 ATCGGGCTATCTTAGAACGTGTGGTATTTTTCAGTCGCCATGCTTGGCTCGCCGAC 152
Db 39 ACCAGAAATTGAAGTTGGAATCTGTGTCATTGCTCCAGACACGCTGTAGAGTCCAAC 98
Qy 153 CAA-ACAAACGACGCTTATGAATGATGTACCCAGATAAGTGGCCACAATGGCCGGTAAA 211
Db 99 TAAGCTACTCAGTTGATGCAAGATGTTACTCCAGATGCTTGGCTACCTGGCTGTAA 158
Qy 212 AGCGGGTATTTAAACGCCACGCTGTGGGAGTTGGTCACTGATGGGGGGTTTATGG 271
Db 159 GTTGGGTGAATTGACTCCAGAGGGTGTGTAATTGCTTACTTGGGTCACTACTGGAG 218
Qy 272 TGATTACTTTCGACGCTTGTGTTTGT--AGCGCGGGATGTCGGCAGAGGGGGGT 328
Db 219 ACAAGATGGTGTGCTGATGGTTTGTTCGCAAGATGTTGTTCCAGATCTGTCAAGT 278
Qy 329 ATATGCACAGGAGATATCGATCAACGPAACCGCTTAAACCGGACAGGATTTCTTGATGG 388
Db 279 TGCTATCATTTGCTGATGTTGATGAAGAACTAGAAAGACTGCTGTAAGCTTCGCTGCCGG 338
Qy 389 TGTTGGCTCGGGGTGGTGGTTCAGCTGCATAATCAGCCCGATTTGAAAGAGACGATCC 448
Db 339 TTTGGCCCACTGCTGCTATCACTGTTCACTCACTCACTCACTCACTCACTCACTCACT 398
Qy 449 CCTGTTCCATCGGTAGAGACTGGCGTGTGTAAGTTAGACAAACGACAAACAGATAAGC 508
Db 399 ATTTGTTCAACCATTTGAAGACTGGTGTCTGTCAATTGGATAACGCTTAACGTTACTGATGC 458
Qy 509 GATTGAAGAACGATTTGGCGGGCGTGTAGATACGGTGAGCCGCTACGCTAAACCTTT 568
Db 459 CATCTGGAAGAGCTGGTGGTTCATCGCTGACTTCTGCTGCTCACTACCAAACTGCCCTT 518

QY 569 TGCCAGATGGGACGTGCTGATTTTGGGCTTCTCCTATTGCAATCTTTGCAACA 628
Db 519 CAGAGAAATGGAAGAGTCCTTTGAACTTCCCAAACTAACTTGTGTTGAAGAGAGAA 578
QY 629 GCAAGGAAACAGTGTGATTTTGGCCACTTTGGCGCAATCAAGTTAACTTAATAAGA 688
Db 579 GCAAGAGCAATCTTGTTCCTTGACTCAAGCCTTGCCATCTGAATTGAAGTCTCTGCTGA 638
QY 689 AGGACAAAGTGAACCTCAGTGGGCACTGGGCTTATCATCGACATGGGTTGAATCTT 748
Db 639 TTG-----TGTCTCTTGAAGTGTGCTGCTCTCTGGCTTCTATGTGTGACTGAAATCTT 692
QY 749 CTTATTACAAACGACAAAGCCATGCGAGAGTTGCTGCAACGGCTAAAGAGGGCGGA 808
Db 693 CTTGTTGCAACAAGCTCAGGTATGCGCAACCAAGTGTGGGTAGATCACTGATCTCA 752
QY 809 GAATGGGTATCTTATGTCATTAATACAGCGCAATTAATTAATGCAAAACACC 868
Db 753 CCAATGGAACACCTTGTGCTTGTGCAACAGCTCAATTCGATTTGTCGAGAGAACTCC 812
QY 869 GTATATCGCCGTCATAAAGGAGCGCCATTATTACAGCAGATAGATACGGCTTTAACCT 928
Db 813 AGAAGTGTCTAGATCCAGAGTACTCCATTGTTGGACTTGATCAAGACCCCTTGAC--- 869
QY 929 TCAACTGGATGCTCAGGGGCAAAAGCTACCCATTTCAGCCCAAAACCGGGTCTTGTCTCT 988
Db 870 TCCACACCCACCAAGCAAGCTTACGGGTGTACCTTGCCAACTTCTGCTTGTTCAT 929
QY 989 CGTGGGCATGATACCAATATGTCATATGCGGTATGCTGGAGCCGACTGGCAGCT 1048
Db 930 TGCCGGTCAAGATCACTTACCTTGGCTTAACTTGGGTGGTCCCTTGAATGGAACCTGACCTT 989
QY 1049 ACCGAGCAACCTGATTAATCTCACCAGGTGGGGATTTGTTTGAACATATGGCAAAA 1108
Db 990 GCCAGGTCAACAGATAAATCACTCCACAGGTGGTGAATTTGCTTTCGAAAGATGGGCTG 1049
QY 1109 TCCGATTAACACAGCGCTACGTTCGGTGAAGATCTTACCAACAGTGAATCACTT 1168
Db 1050 ACTGTCGATTAATCTCAATGGAATCAAGTCTCCTTGGTCTTCCAAACCTTGCACAAAT 1109
QY 1169 GCGTAATCCGAAATTTGGATCTGAAAAATAACCCAGCGGTATTTATTTCCGTTGCGAT 1228
Db 1110 GAGAGACAAGACTCCATTGCTTCTTGAACACTCCACAGGTGAAGTCAAGTTGACCTTGGC 1169
QY 1229 TGCTGTTGTGAAATACGTTGACGATAGCTTTGGAGTTGATACATCCAAA 1284
Db 1170 TGGTTGTGAAAGAAACGCTCAAGGTATGTGTTCTTTGGCTGGTTTCACTCAA 1225

RESULT 6

AAC68299
ID AAC68299 standard; DNA; 5421 BP.

XX
AC AAC68299;
XX

DT 20-FEB-2001 (first entry)

XX SV40/APPA plasmid coding sequence.

DE Transgenic animal; salivary protein; phytase; phosphorus; animal growth;
KW environmental pollution; pig; ds.

XX Rhesus macaque polyoma virus.

OS Escherichia coli.

OS Chimeric.

XX WO200064247-A1.

XX 02-NOV-2000.

XX 20-APR-2000; 2000WO-CA000430.

XX 23-APR-1999; 99US-0130508P.

XX (UYGU-) UNIV GUELPH.
XX Forsberg CW, Golovan S, Phillips JP;
XX WPI: 2000-687245/67.
DR P-PSDB; AAB36262.
XX Transgenic non-human animal for gastrointestinal tract specific
PT expression of a protein, preferably phytase, comprises a nucleic acid
PT sequence including a heterologous transgene construct encoding the
PT protein.
XX Claim 56; Fig 22; 152pp; English.
XX The present invention provides transgenic animals which produce desired
CC proteins, in this case pigs which expresses phytase in the salivary
CC gland. Low phytase production levels result in phytate in the diet being
CC excreted and causing phosphorus contamination in water, as well as
CC reducing the growth of animals. The invention provides a number of
CC transgenes containing the E. coli APPA phytase coding sequence
XX
SQ Sequence 5421 BP; 1413 A; 1321 C; 1331 G; 1355 T; 0 U; 1 Other;
Query Match 17.1%; Score 226.4; DB 3; Length 5421;
Best Local Similarity 53.3%; Pred. No. 2e-62;
Matches 572; Conservative 0; Mismatches 487; Indels 14; Gaps 4;
QY 107 TAGAACGTGTGCTTATTTTGTGTCGCATGCTGCTCGCGACCAA-ACAACGACGC 165
Db 128 TGGAAAGTGTGCTGATTTGCTGCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 187
QY 166 TTATGAATGATGTTACGCGCAGATAAGTGGCCAAATGCGCGGTAAAGCGGGTATTTAA 225
Db 188 TGATGACGATGTCACCCACAGACGATGGCAACCTGGCCGCTAAACCTGGGTTGGCTGA 247
QY 226 CGCCAGGTGTGCGGAGTGTGTCATTTGATGGGGGGGTTTTATGCTGATTTTCGCA 285
Db 248 CACCGCGNGTGTGAGCTAATTCGCTATCTCGACATTAACACGCGCAGCTCTGGTAG 307
QY 286 GCGTTTGTGTTTGTAGCG---GCGGATGTCGCGCAGAGGGGGGTATATGCACAGGAG 342
Db 308 CCGACGATTTGTGCGGAAAGAGGCTGCCGCGAGTCTGGTCAGTCCGATTTATGCTG 367
QY 343 ATATGATCAAGTACCGCTTAACCGACAGGCAATTTCTTGATGCTGCTGCTGCTGCTG 402
Db 368 ATGTCGACGAGCTACCGCTAAACACAGCGGAGCCCTTCGCGCGGGGTGGCAGCTGACT 427
QY 403 GTGGTTTGACCGTGCATTAATCAGGCGGATTTGAAAAAGACCGATCCCTGTTTCCATCCGG 462
Db 428 GTGCAATAACCGTACATACCGAGGAGATACGTCAGTCCGATCCGTTATTTATCTCTC 487
QY 463 TAGACATGCGGTGTGTAAGTTAGAACACGCAACAGATAAGCGGATTTGAGAACGAT 522
Db 488 TAAAAACTGGCGTTTGCCTGCACTGGATAACGCGAACGTGACGCGATCTCTCAGCAGGG 547
QY 523 TGGCGGGCGGTTAGATACGGTGAAGCCAGCGCTACGCTAAACCTTTTCCAGATGGGG 582
Db 548 CAGGAGGCTCAATTCGCTGACTTACCGGCGATCGCAACCGCGTTTCGCGACTGGAAC 607
QY 583 ACGTGTGAATTTTCGCGCTTCTCTTATGCAATCTTTGCAACAGCAGAGGAAACGTT 642
Db 608 GGGTGTCTTAATTTTCCGCAATCAAACTTGTGCTTTAAACGTCGAGAAACAGGACGAAAGCT 667
QY 643 GTGATTTTCCCACTTTGGGCGCAATGAAGTTAAGCTTAATAAAGAACGACAAAAGTGA 702
Db 668 GTTCATTAACGCGGATTAACATCGGAATCTGAGTGAAGCGCGA-----CAATGCT 721
QY 703 CCTCAGTGGGCACTGGCGTTTATCATCGACATTTGGTGAATCTTCTTTATTTACAAAACG 762
Db 722 CATTAACCGGTGGGTGAGCTCGCATCAATGCTGACGAGATATTTCTCTGCAACAG 781
QY 763 CACAGCCATGCCAGGTTGCTTGGCAACCGCTAAAGGGGGGAGGATTTGGGTATCTCT 822

Db 782 CACAGGAAATGCGGAGCCGGGTGGGAAGGATCACCGATTACACCAAGTGGAAACACT 841
Qy 823 TATGTGATTAACATACGCGCAATTTAAATTTATGCGAAACACCGTATATCGCCGTC 882
Db 842 TGTAGTTTGCATACGCGCAATTTATTTGCTACACGCGCAGAGGTTGCCGCA 901
Qy 883 ATAAAGGAGCGCATTTATACAGCAGATAGATACGCTTTAACTTCACTGATGCTC 942
Db 902 GCGCGCCACCCCGTTTATTAGATTGATCAAGACAGCGTTGACGCGCCACCGCAAAA 961
Qy 943 AGGGGCAAAAGCTACCATTTACGCCCAAAACCGGGTCTTTCTCGGTGGCATGATA 1002
Db 962 ACAGGCGTATGGTGCATATACCACTTCA----GTGCTGTTATCGCGGACAGATA 1017
Qy 1003 CCAATATGCGCAATATGCGGTATGCTGGAGCGCACTGGCAGCTACCGAGCAACCTG 1062
Db 1018 CTAATCTGGCAATCTCGCGCGCACTGGAGCTCAACTGGAGCTTCCCGGTACGCGG 1077
Qy 1063 ATAATCTCCACAGGTGGGGATTGGTTTGAACATATGGCAAAATCCGATACCAACC 1122
Db 1078 ATAAACGCGCGCAGGTGGTGAAGTGTGTTGAACGCTGGGTGCGCTAAGCGATACA 1137
Qy 1123 AGCGTACGTTGCGGTGAAGATTTTACAAACAGATGGATCAGTTGCGTAAT 1175
Db 1138 GCCAGTGGATTACAGTTTGGCTGCTCCAGACTTTACAGCAGATGCGTGAT 1190

RESULT 7

ADC87742

ID ADC87742 standard; DNA; 1308 BP.

XX AC ADC87742;

XX DT 01-JAN-2004 (first entry)

XX DE DNA encoding the modified Escherichia coli phytase.

XX KW Phytase; food supplement; enzyme delivery matrix; soybean meal;

XX KM thermotolerance; thermostability; kernel; phytate;

XX KW myo-inositol-hexaphosphate; inositol; inorganic phosphate;

XX KW thermotolerant; feed value; digestion; gene; ds.

XX OS Synthetic.

OS Escherichia coli.

XX FH Key Location/Qualifiers

XX CDS 1..1308

XX FT /*tag= a

XX FT /product= "Modified phytase"

XX PN US2003103958-A1.

XX PD 05-JUN-2003.

XX PF 24-MAY-2002; 2002US-00156660.

XX PR 13-AUG-1997; 97US-00910798.

XX PR 01-MAR-1999; 99US-00259214.

XX PR 13-APR-1999; 99US-00291931.

XX PR 25-MAY-1999; 99US-00318528.

XX PR 25-MAY-2000; 2000US-00580515.

XX PR 24-MAY-2001; 2001US-00866379.

XX PA (DIVE-) DIVERSA CORP.

XX PI Short JM, Kretz K, Gray KA, Barton NR, Garrett JB, O'donoghue E;

XX PI Mathur EJ;

XX XX WPI; 2003-787039/74.

XX DR P-PSDB; ADC87743.

XX PI New nucleic acid encoding a polypeptide having phytase activity, useful

PT in improving the feeding value of phytate rich ingredients or as an aid
PT in phytate digestion.
XX Claim 9; SEQ ID NO 1; 113pp; English.
XX The invention discloses a new isolated or recombinant nucleic acid which
CC encodes a polypeptide having a phytase activity. Also claimed is a
CC nucleic acid probe, an amplification primer sequence pair, an expression
CC cassette comprising the nucleic acid, a vector comprising the nucleic
CC acid, a transgenic non-human animal or plant, or its seed, comprising the
CC nucleic acid, an antisense oligonucleotide, inhibiting the translation of
CC a phytase message in a cell, a heterodimer comprising the polypeptide and
CC a second domain, an array comprising immobilised polypeptide or nucleic
CC acid, a hybridoma comprising an antibody that specifically binds to the
CC polypeptide, a food supplement for an animal, an edible enzyme delivery
CC matrix, an edible pellet comprising a granule edible carrier and the
CC polypeptide, a feed composition, a soybean meal, isolating or identifying
CC the polypeptide, making an anti-phytase antibody, producing a recombinant
CC polypeptide, determining whether a compound binds to the polypeptide,
CC identifying a modulator, whole cell engineering of new or modified
CC phenotypes by using real-time metabolic flux analysis, increasing
CC thermotolerance or thermostability of the phytase polypeptide, increasing
CC the resistance of the phytase polypeptide to enzymatic inactivation in a
CC digestive system and processing of corn and sorghum kernels. The phytase
CC activity comprises catalysis of phytate (myo-inositol-hexaphosphate) to
CC inositol and inorganic phosphate and is thermotolerant. The nucleic acid
CC is useful in improving the feeding value of phytate rich ingredients or
CC as an aid in phytate digestion. The sequence presented is the DNA
CC encoding the modified Escherichia coli phytase.
XX SQ Sequence 1308 BP; 315 A; 345 C; 357 G; 291 T; 0 U; 0 Other;
Query Match 17.1%; Score 226.2; DB 9; Length 1308;
Best Local Similarity 53.3%; Pred. No. 1e-62;
Matches 572; Conservative 0; Mismatches 488; Indels 13; Gaps 4;
Qy 107 TAGAACGTGGTTATTTTGGTCCGATGTTGCTCGCGACCAA-ACAACGAGC 165
Db 89 TGGAAAGTGGTGAATTTGTCAGTGTGCTGCTCCACAGGACGCGAC 148
Qy 166 TTATGAATGATTTACCGCAGATAAGTGGCCACAATGCGGTAAGAGCGGGTATTAA 225
Db 149 TGATGAGGATGTCAACCCAGACGATGGCCCACTGGCGGTAACCTGGTGAGCTGA 208
Qy 226 CGCCAGTGGTGGAGTTGGTCAATGATGGGGGTTTATGGTGATTTACTTTCGCA 285
Db 209 CACCGCGGTGGTGAAGCTAATCGCCTATCTCGGACATTAAGTGGTGGTGGTAG 268
Qy 286 GCCTTGGTTTGTAGC---GGCGGGATGTCCGCGAGAGGGGGGTATATGCAAGGCAG 342
Db 269 CCGACGGATTGCTCCTAAATGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 328
Qy 343 ATATCGATCAACGTTACCGCTTAACCGCAGAGCAATTTCTGTGTTGGTGGTGGTGGT 402
Db 329 ATGTGCGAGCGTACCGCTTAACCGCAGAGCGCTTTGCGCGGGGTGGCCTGACT 388
Qy 403 GTGGTTTGACCGTGCATAATCAGGCGCATTTGAAAAAGACCGATCCCTGTTCATCCGG 462
Db 389 GTGCAATAACCGTACATACCGCAGAGATAGTCCAGTCCCGATCCGTTATTTAATCCTC 448
Qy 463 TAGAGACTGGCGTGTGAAGTTAGAACGCAACGCAACAGATTAAGCAAGATGAAGACAT 522
Db 449 TAAAACTGGGTTTGGCACTGATTAACGCGAGCGTGGTGGTGGTGGTGGTGGTGGTGG 508
Qy 523 TGGGCGGGCGCTTAGATACGCTGAGCGAGCGCTAGCTAAACCTTTTGGCCAGATGGGG 582
Db 509 CAGGAGGTCAATTTGCTGACTTTACCGGCGATTAATCAACGCGCTTTTCGGAATGGAC 568
Qy 583 ACGTGTCAATTTTGGCGCTTCTCCTTATTGCAAAATCTTTTGCACACAGCAAGAAAAACGT 642
Db 569 GGGTGTCTTAATTTTCGCAATCAAACTGTGCTTTAAACGCTGAGAAACAGGACGAAAGCT 628
Qy 643 GTGATTTGCCCACTTTTGGCGCCCAATGAGTTAAGTTAAGTAAAGAGGACAAAGTGA 702

Db 629 GTTCATTAAACGAGGATTAACCATCGGAATCAAGGTGAGCGGCACTGTCTCTATTAA 688
Qy 703 CCCTCAGTGGGCACTGCGGTTATCATCGACATTTGGGTGAATCTTCTATTACAAACG 762
Db 689 CC-----GTTGGGTAGCTTCGATCAATGCTGACGGAGATTTCTCTCTGCAACAG 742
Qy 763 CACAAGCCATGCGAGGTTGCTGCGCAACGGCTAAAGGGGCGGAGATTTGGTATCCT 822
Db 743 CACAGGGAATGCGGAGCGGCGGTGGGAAGGATCACCGATTACACCGTGGAAACCT 802
Qy 823 TATTCTCATTAACAACGCGAATTTAATTAATGGCAAAACACCGTATATCGCCGTC 882
Db 803 TGCTAAGTTTGCATTAACCGGCAATTTGATTGCTACACGCGCAGAGGTTCGCCGCA 862
Qy 883 ATAAAGGACGCGATTAATACAGCAGATAGATAGCGGCTTTAACCTTCAACTGGATGCT 942
Db 863 GCGCGCCACCGGTTATTAGATTTGATCAAGACAGCGTTGACGCCCATC---CACCGC 919
Qy 943 AGGGCCAAAGCTACCCATTTACGCCCAACCGGCTTCTCTCGTGGGATGATA 1002
Db 920 AAAAAACGCGTATGATGATTAACCATTTCCACTTTCAGTCTGTTATCCCGGACAGATA 979
Qy 1003 CCAATATTGCCAATATTGGGATGCTGGGAGCGGACTGGCAGCTACCCGAGCAACCTG 1062
Db 980 CTAATCTGGCAATCTCGCGCGCACTGGAGCTCACTGGACGCTTCCCGGTGAGCGG 1039
Qy 1063 ATATATCTCCACGAGTGGGATTTGGTTTGTACTATGCAAAATCCGATACCAACC 1122
Db 1040 ATACACGCGCGAGTGGTGAATCGGTGTTGAACGCTGCGTGGCTGAAGCGATAACA 1099
Qy 1123 AGCGTACGTTGCGGTGAAGATGTTCTACCAAAACGATGGATCAGTTGGGTAA 1175
Db 1100 GCCAGTGGATTACAGTTTCGCTGGTCTTCCAGACTTTACAGCAGATGCGTGT 1152

RESULT 8

AD25463
ID AAD25463 standard; DNA; 1901 BP.
XX AC AAD25463;
XX DT 26-MAR-2002 (first entry)
XX DE Escherichia coli appA phytase wild type DNA.
XX KW Bacterial phytase; Kl2 appA phytase; protease stability; anabolic;
KW gastrointestinal; nutritional value; feed treatment process; therapy;
KW thermal tolerance; growth performance; alcoholic drink; biopulping;
KW non-alcoholic drink; biobleaching; ds.
XX OS Escherichia coli.
XX FH Key Location/Qualifiers
XX CDS 188..1486
FT /tag= a
FT /product= "E. coli appA phytase protein"
XX PN WO200190333-A2.
XX PD 29-NOV-2001.
XX PF 24-MAY-2001; 2001WO-US017118.
XX PR 25-MAY-2000; 2000US-00580515.
XX PA (DIVE-) DIVERSA CORP.
XX PI Short JM, Kretz KA, Gray KA, Barton NR, Garrett JB, O'donoghue E;
XX WPI; 2002-083108/11.
XX DR P-PSDB; AAE15807.
XX

PT New bacterial phytase for e.g. improving the nutritional value of phytate
PT -containing foodstuffs and subsequently improving the growth performance
PT of an organism that consumes it, or in treating animal digestive systems.
XX Claim 42; Fig 7; 170pp; English.
PS
XX The patent discloses recombinant bacterial phytase from Escherichia coli
CC Kl2 appA phytase. The enzyme has phytase activity and improved thermal
CC tolerance when compared with wild-type phytase. It has improved protease
CC stability at low pH. The recombinant phytase is useful for improving the
CC nutritional value of phytate-containing foodstuffs and subsequently
CC improving the growth performance of an organism that consumes it, in
CC treating animal digestive systems, in feed treatment processes and for in
CC vitro purposes related to research, discovery and development. They are
CC also used for generating recombinant digestive system life forms, for
CC producing or manufacturing alcoholic and non-alcoholic drinks based on
CC the use of moulds, grains and/or plants, in biopulping and bio-bleaching
CC where a reduction in the use of environmentally harmful chemicals that
CC are traditionally used in the pulp and paper industry is desired and in
CC the reduction or possible elimination of the need for mineral
CC supplements, enzymes or therapeutic drugs for animals from the daily feed
CC thus increasing the amount calories and nutrients present in the feed.
CC The present sequence is a DNA encoding E. coli appA phytase wild type
CC protein
XX
SQ Sequence 1901 BP; 474 A; 499 C; 499 G; 428 T; 0 U; 1 Other;

Query Match 17.0%; Score 225.2; DB 6; Length 1901;
Best Local Similarity 53.2%; Pred. No. 2.7e-62;
Matches 571; Conservative 0; Mismatches 489; Indels 13; Gaps 4;
Qy 107 TAGAAGCTGTGGTATTATTTGAGTCGCCATGCTGTTTCGCTCGCGCAGCAA-ACAACGCAAG 165
Db 276 TGGAAAGTGTGGTATTGTCAGTCGTCATGTTGCTGCTCCAACCAAGGCCACGCAAC 335
Qy 166 TTATGAATGATGTTACGCCAGATAGTGGCCCAATATGGCCGGTAAAGCGGGGTATTAA 225
Db 336 TGATGCAGGATGTCACCCAGACGCAATGGCCCACTTGGCCGGTAAACTGGGTGCTGA 395
Qy 226 CGCCACGCTGTGGGAGTGTGTCACATTTGATGGGGGGTATTATGCTGATTAATTTCGCA 285
Db 396 CACCGCGGTGTGGTGAATATCGCTATCTCGACATTAACCAACGCCAGCTCTGGTAG 455
Qy 286 GCCTTGGTTTGTAGCG---GCGGATGTCGCGCAGAGGGGGGGGTATATGCAAGCGAG 342
Db 456 CCGACGATTTCTGGCGGAAAAGGGTGGCCGAGTCTGTCTGTCAGGTCTGGGATTTTGTG 515
Qy 343 ATATCGATCAACCTTACCGCTTAACCGCAGCAGCATTTCTTGATGTGTGGTCCCGGGT 402
Db 516 ATGTGACGAGCGTACCCGTTAAACAGGCGAGCCTTCGCCCGCGGGCTGGACCTGACT 575
Qy 403 GTGGTTTGACCGTGCATAATCAGGCGGATTTGAAAAGACCGATCCCTTTCATCCCGG 462
Db 576 GTGCAATAACCGTACATACCCAGGCAGATACGTCAGTCCCGATCCGTTATTATATCCTC 635
Qy 463 TAGAGACTGCGTGTGTAGTAGACACGCAACACAGATTAAGCGATTGAAGACGAT 522
Db 636 TAAAACTGGCGTTTGGCCAACTGGATAACGCGAACGCTGACTGACCGGCTCTCAGCAGG 695
Qy 523 TGGCGCGCGCTTAGATACGCTGAGCGCGCTACGCTAAACCTTTTGGCCAGATCGGGG 582
Db 696 CAGGAGGTCAATTGCTGACTTTACCGGCATCGGCAACCGGCTTTCGCGAATCGAAC 755
Qy 583 ACGTGCTGAATTTTGGGCTTCTCTTATGCAATCTTTGCAACAGCAGGAAACCT 642
Db 756 GGGTGTCTTAATTTTCCGCAATCAAACTGTGCTTAAACGCTGAGAACAGGACGAAGCT 815
Qy 643 GTGATTTTGGCCACTTTTGGCGCAATGAAGTTAACTGTTAATAAGAGGACAAAAGTGA 702
Db 816 GTTCATTAACGAGGCACTTACCATCGGAATCAGGTGAGCGCGCA-----CAATGTCT 869
Qy 703 CCCTCAGTGGGCACTGGCGTTATCATTCGACATGGGTGAAATCTTCTTTATCAAAACG 762

E. coli K12 appA gene encoding phytase.
phytase; ds; gene; phytate; appA gene; animal feed; inorganic phosphate;
digestion enhancement; transgenic; thermal tolerance; protease stability.
Escherichia coli; strain K12.

US2002136754-A1.

26-SEP-2002.

24-MAY-2001:

13-AUG-1997: 97US-00910798

01-MAR-1999; 99US-00259214.

25-MAY-1999; 99US-00318528.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841

(SHUR/) SHORT J. M.
(KRET/) KRETZ K. A.

(GRAY//) GRAY K A.
(RABT//) RABTON N P

(GARR/) GARRETT J B.

(MATH/) MATHUR B J.

Short JM, Kretz KA, Gray KA

U' Donoghue E, Mathur EU;

WPI; 2003-040002/03.
D-DESB. 200318450

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

enzyme, useful in the product

functional value of phyceae digestion in humans and animals

Alajm 13. Eic 7. Eom. Enljo

100

encoding a phytase enzyme app

Also included the *E. coli* app

derived from it) or its mutant host cells. a method of impro

containing foodstuff by conta-

the liberation of inorganic p

microbial phytase (comprising

plant under conditions where converting the plant cells, p

reed), a feed composition for cells, plant parts or plants

foodstuff), a method to treat
digestive enhancement by the

comprising administering to t

contain an expression system

encoding a pyruvate kinase, a comprising a heterologous nucleic acid sequence encoding a pyruvate kinase, a

having phytase activity. The nutritional value of phytate-

animal feed and for enhancing

also improves the feeding val-

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

RESULT 10
ADA19449
ID ADA1
XX
AC ADA1
XX
DT 20-N
XX

XX	Sequence	1901 BP; 474 A; 499 C; 499 G; 428 T; 0 U; 1 Other;
SQ	Query Match	17.0%; Score 225.2; DB 7; Length 1901;
	Best Local Similarity	53.2%; Pred. No. 2.7e-62;
	Matches 571; Conservative	0; Mismatches 489; Indels 13; Gaps 4;
QY	107	TAGAACGTGTGGTATTTTTCAGTCGCGCATGGTGTTCGCTCGCGACCAAA-ACAACGACG 165
DB	276	TGGAAGTGTGGTATTTTCAGTCGCGCATGGTGTTCGCTCGCGACCAAA-ACAACGACG 335
QY	166	TTATGAATGATGTTTACGCCAGATAAGTGGCCACAATGGCGGTAAAGGGGGTATTTAA 225
DB	336	TGATGAGGATGTCACCCAGAGCATGGCCAACTGGCGGTAAAGGGGGTATTTAA 395
QY	226	CGCCAGTGTGCGGATGGTGCATATGATGGGGGGGTTTATGATGATTAATCTTCGCA 285
DB	396	CACCGGNGGTGTGAGTAATCGCTATCTCGCAATATACCAACGCCAGCGCTGTGAG 455
QY	286	GCCTTGTGTTTATAGG---GCGGGATGTCGCGACAGGGGGGGTATATGACAGGCGAG 342
DB	456	CGACGATGTCGGGGGAAAGGGGTCGCCGAGTCTGTCAGTCGCGATATTGCTG 515
QY	343	ATATCGATCAACGTAACCGGTTTAAACCGGACAGGCATTTCTGATGTGTGGCTCCGGGT 402
DB	516	ATGTCACGAGCGTACCGGTAAACAGGCGAAGCCTTCGCGCGCGCTGGCACCTGACT 575
QY	403	GTGTTTGACGTCATATCAGCGCATTTGAAAGAGACCGATCCCTGTTCCATCCG 462
DB	576	GTGCAATACCGTATACCCAGCGAGATACGTCCAGTCCGATCCGTTATTTAATCTTC 635
QY	463	TAGAGATGCGGTGTGTAAGTTAGACAACGCAACAAACAGATAAAGCGATTTGAAGAACGAT 522
DB	636	TAAAACTGCGGTTTCCAACTGATAACGCGAAGCTGACTGACGGATCTCTCAGCAGG 695
QY	523	TGGCGCGCGTTAGATACGGTACGGTACCGCTACGCTAAACCTTTGCGCCAGATGGGG 582
DB	696	CAGAGGGTCAATGCTGACTTTACCGGGATCGGCAACCGGCTTCGCGACCTGGAAC 755
QY	583	ACGTGCTGAAATTTGCGGCTTCTCTTATTTGCAAACTTTTGCAACAGCAGGAAACCT 642
DB	756	GGGTGCTTAATTTTCCGCAATCAACTTGTGCTTAAACGTGAGAAACAGGACGAAAGCT 815
QY	643	GTGATTTTGGCCATTTGGCGCAATGAAGTTAACGTTAATGAAGAGGACAAAGTGA 702
DB	816	GTTCAATTAACGCGAGCATTAACCATCGAATCAAGGTGAGCGCGA-----CAATGCT 869
QY	703	CCCTCAGTGGCCCACTGGCGTTATCATCGACATTTGGTGAAATCTTTTATTACAAAAG 762
DB	870	CATTACCGGTGGGTAAGCTCGCATCATGCTGACGGAGATATTCTCTCTGCAACAG 929
QY	763	CACAGCCATGCGAGAGTTGCTCGCAACGGCTAAAGGGGCGGAGATTTGGGTATCT 822
DB	930	CACAGGGAATGCGCGAGCGCGGGTGGGGAAGSATCCCGATTACACCCAGTGGAAACCT 989
QY	823	TATTGTCAATACATACGCGCAATTTAATTTAATGGCAAAACACCGTATATCGCCCGTC 882
DB	990	TGCTAAGTTTGCATAACCGCAATTTATTGTCTACACGCGCAGAGGTGCGCGCA 1049
QY	883	ATAAAGGAGCGCATTTATACAGCAGATAGATAGCGGTTTAAACCTTCACTGAGTGTCT 942
DB	1050	GCGCGCCACCGCGTTATTAGATTGTATCAAGACAGCGTTTGAAGCGCCCATC---CACCG 1106
QY	943	AGGGGCAAAAGTACCCATTTACGCCCAAAACCGGCTTGTTCCTCGGTGGCGCATGATA 1002
DB	1107	AAAAAGGGGTATGTTGTGACATTACCATCTTCACTGTTGTTATGCGCGGACAGATA 1166
QY	1003	CCATATTGCAATTTTTCGGGTATGCTGGAGCCGACTGGCAGCTACCGGACCAACCTG 1062
DB	1167	CTAATCTGGCAAACTCTCGCGGGCGCACTGGAGTCACTGAGAGCTTCCTCGGTGAGCGG 1226
QY	1063	ATATACTCCACAGGTGGGGATGTTGTTTTCGACTATGCGCAAAATCCGATACCAACC 1122

Db	1227	ATACACGCCGCCAGGTGGTGAACCTGGTGTGAAAGCTGGCTGGCTGGCTGAAGGATAACA 1286
QY	1123	AGCGCTACGTTTGGGTGAGATGTTCTACCAACAGATGATGATGATGATGATGATGAT 1175
Db	1287	GCCAGTGGATTTCAGTTTCGCTGCTCTCCAGACTTTACAGCAGATGCGGTGAT 1339
RESULT 11		
ADC87744		
ID	ADC87744	standard; DNA; 1901 BP.
XX	AC	ADC87744;
XX	DT	01-JAN-2004 (first entry)
XX	DE	DNA encoding Escherichia coli appA phytase.
XX	KW	Phytase; food supplement; enzyme delivery matrix; soybean meal;
XX	KW	thermotolerance; thermostability; kernel; phytate;
XX	KW	myo-inositol-hexaphosphate; inositol; inorganic phosphate;
XX	KW	thermotolerant; feed value; digestion; gene; ds; appA.
XX	OS	Escherichia coli.
XX	Key	Location/Qualifiers
XX	FT	188..1486
XX	FT	/*tag= a
XX	FT	/product= "AppA phytase"
XX	FT	/transl_except= (pos:401..403,aa:Arg)
XX	PN	US2003103958-A1.
XX	PD	05-JUN-2003.
XX	PF	24-MAY-2002; 2002US-00156660.
XX	PR	13-AUG-1997; 97US-00910798.
XX	PR	01-MAR-1999; 99US-00259214.
XX	PR	13-APR-1999; 99US-00291931.
XX	PR	25-MAY-1999; 99US-00318528.
XX	PR	25-MAY-2000; 2000US-00580515.
XX	PR	24-MAY-2001; 2001US-00866379.
XX	PA	(DIVE-) DIVERSA CORP.
XX	PI	Short JM, Kretz K, Gray KA, Barton NR, Garrett JB, O'donoghue E;
XX	PI	Mathur EJ;
XX	DR	WPI; 2003-787039/74.
XX	DR	P-PSDB; ADC87745.
XX	PT	New nucleic acid encoding a polypeptide having phytase activity, useful
XX	PT	in improving the feeding value of phytate rich ingredients or as an aid
XX	XX	in phytate digestion.
XX	PS	Example 1; SEQ ID NO 3; 113pp; English.
XX	CC	The invention discloses a new isolated or recombinant nucleic acid which
XX	CC	encodes a polypeptide having a phytase activity. Also claimed is a
XX	CC	nucleic acid probe, an amplification primer sequence pair, an expression
XX	CC	cassette comprising the nucleic acid, a vector comprising the nucleic
XX	CC	acid, a transgenic non-human animal or plant, or its seed, comprising the
XX	CC	nucleic acid, an antisense oligonucleotide, inhibiting the translation of
XX	CC	a phytase message in a cell, a heterodimer comprising the polypeptide and
XX	CC	a second domain, an array comprising immobilised polypeptide or nucleic
XX	CC	acid, a hybridoma comprising an antibody that specifically binds to the
XX	CC	polypeptide, a food supplement for an animal, an edible enzyme delivery
XX	CC	matrix, an edible pellet comprising a granule edible carrier and the
XX	CC	polypeptide, a feed composition, a soybean meal, isolating or identifying
XX	CC	the polypeptide, making an anti-phytase antibody, producing a recombinant
XX	CC	polypeptide, determining whether a compound binds to the polypeptide,
XX	CC	identifying a modulator, whole cell engineering of new or modified
XX	CC	phenotypes by using real-time metabolic flux analysis, increasing

CC thermotolerance or thermostability of the phytase polypeptide, increasing
CC the resistance of the phytase polypeptide to enzymatic inactivation in a
CC digestive system and processing of corn and sorghum kernels. The phytase
CC activity comprises catalysis of phytate (myo-inositol-hexaphosphate) to
CC inositol and inorganic phosphate and is thermostolerant. The nucleic acid
CC is useful in improving the feeding value of phytate rich ingredients or
CC as an aid in phytate digestion. The sequence presented is the DNA
CC encoding the *Escherichia coli* appA phytase.
XX
SQ Sequence 1901 BP; 474 A; 499 C; 499 G; 428 T; 0 U; 1 Other;
Query Match 17.0%; Score 225.2; DB 9; Length 1901;
Best Local Similarity 53.2%; Pred. No. 2.7e-62;
Matches 571; Conservative 0; Mismatches 489; Indels 13; Gaps 4;
Qy 107 TAGAACGTGGTATTATTTAGTCGCATCGTGTGCTCGCCGACCAA-ACAACGACG 165
Db 276 TGGAAAGTGGTGATGTCAGTCGTGTCATGCTGTGCTGCTCCAAACCAAGCCACGCAAC 335
Qy 166 TTATGATGATGTTACCCAGATGAAGTGGCCACATGCGCCGTAAGAGCGGGTATTAA 225
Db 336 TGATGAGGATGTCACCCAGAGCATGGCCACCTGGCCGCTAAACTGGTGGCTGA 395
Qy 226 CGCCACGTGGTGGAGTTGGTCAATGATGGGGGGTTTTATGGTGAATTAATTTCGCA 285
Db 396 CACCGCGGTGGTGAGCTAATCGCTATCTCGACATTAACCAACGCGCTCGTAG 455
Qy 286 GCCTGGTTGTTAGG---GCGGATGTCGGCAGAGGGGGGGTATATGCACAGCGAG 342
Db 456 CGGACGGAATGCTGGCGAAAGAGGCTGCCCGCAGTCTGTGTCAGGTGCGGATTAATGCTG 515
Qy 343 ATATCGATCAACGTACCGGCTTAAACCGGACAGGCAATTTCTTGATGGTGGTCCCGGGT 402
Db 516 ATGTCGACGAGCGTACCGGTAACACAGCGGAAGCTTCGCGCGGCTGGCACCTGACT 575
Qy 403 GTGGTTTGCAGGTCATTAATCAGGCCGATTTGAAAGACCGATCCCTGTTTCATCCGG 462
Db 576 GTGCAATTAACGTCATACCCAGGACAGATACGTCACGTCGCCGATTCGGTTATTAATCCTC 635
Qy 463 TAGAGACTGGCGTGTGAAGTTAGACAAACGACAAACAGATTAAGCGATTGAAGAACGAT 522
Db 636 TAAAACTGGGTTTGCACACTGATTAACGCGACGTCGACTGACGCGATCTCAGCAGGG 695
Qy 523 TGGCGCGCGTTAGATACGGTAGCGCAGCGCTACGCTAAACCTTTTGGCCAGATGGGG 582
Db 696 CAGGAGGGTCAATTTGCTGACTTTACCGCGGATCGGCAACGGGCTTCGCGAACTGGAAC 755
Qy 583 ACGTGTGTAATTTTGGCGCTTCTCTTATTGCAATCTTTGCAACGACGAAGAAACGT 642
Db 756 GGGTCTTAATTTCCGCAATCAAACTGTGCTTAAACGTGAGAAACAGACGAAAGCT 815
Qy 643 GTGATTTGGCCACTTTGCGCCCAATGAAGTTAACTTAATAAGAGAGGACAAAAGTGA 702
Db 816 GTTCATTAACGACGACATTAACATCGGAACCTCAAGGTGAGCGCGA-----CAATGTCT 869
Qy 703 CCCTCAGTGGCCACTGGCTTATCATCGACATTTGGTGAATCTTCTTATTACAAACG 762
Db 870 CATTAACGGTGGGTGAAGCTTCGATCAATGCTGACGAGATATTCTCTCTGCAACAG 929
Qy 763 CACAAGCCATGCCAGAGTTTCCCTGGCAACGGCTTAAAGGGGGCGAGAAATGGGTATCCCT 822
Db 930 CACAGGGAATCCCGAGCGCGGTGGGGAAGATCACCGATTACACAGCTGGAACACCT 989
Qy 823 TATTGTCAATACATACCGCAATTTAATTAATGGCAAAAACACCGTATATCGCCGCT 882
Db 990 TGCTAAGTTTGCATAACGCGCAATTTTATTTTGTACAAACGACGCGAGAGGTTGCCGCA 1049
Qy 883 ATAAGGAGACGCCATTATTACAGCAGATAGATAGCGGTTTAAACCTTTCACTGGATGCTC 942
Db 1050 GCGCGCCACCCGTTATTAGATTGATCAGACAGCGTTGACGCCCATC---CACCGC 1106
Qy 943 AGGGGCAAAAGCTACCCATTTACGCCCAAACCGGGTCTTGTCTCTCGTGGGCAATGATA 1002

Db 1107 AAAAACAGCGGTATGGTGTGACATTACCACTTCAGTGTCTGTTTATCGCCGACACGATA 1166
Qy 1003 CCAATATTGCCAATATTTCCGGGTATGCTGGAGCGGACTGGCAGTACCCGAGCAACCTG 1062
Db 1167 CTATCTCGCAAAATCTCGCGCGGCACTGGAGCTCAACTGGACGCTTCCCGGTGAGCGG 1226
Qy 1063 ATATACTCCACCAGGTGGGGGATTTGTTTGAATGACTATGGCAAAATCCGGATACACACC 1122
Db 1227 ATACACACCGCCGCGAGGTGGAACCTGGTGTGTTGAACGCTGGCGTGGGTAAAGCGATAACA 1286
Qy 1123 AGCGCTAGCTGGCGTGAAGATGTTCTACCAAAAGATGGATCAGTTGGTGAAT 1175
Db 1287 GCCAGTGATTCAGGTTTCGCTGCTTCCAGACTTTACAGCAGATGCGTAT 1339
RESULT 12
ABK12514
ID ABK12514 standard; DNA; 1299 BP.
XX
AC ABK12514;
XX
DT 05-JUN-2002 (first entry)
XX
DE DNA encoding phytase associated protein.
XX
KW Phytase; gene; ds.
XX
OS Unidentified.
XX
FH Key Location/Qualifiers
CDS 1..1299
FT /*tag= a
FT /product= "Phytase associated protein"
XX
PN KR99086028-A.
XX
PD 15-DEC-1999.
XX
PF 25-MAY-1998; 98KR-00018810.
XX
PR 25-MAY-1998; 98KR-00018810.
XX
PA (WOJ-) WOJIN CO LTD.
XX
PI Bae HD, Forceburgh CW, Goloben S, Cheng KJ;
XX
XX WPI: 2000-645078/62.
DR P-PSDB; AAU77775.
XX
PI Novel phytase gene, recombinant phytase and usage thereof.
XX
PS Claim 1; Fig 2; 10pp; Korean.
XX
CC The invention relates to a novel phytase gene, a recombinant phytase gene
CC and their uses. This sequence encodes a phytase associated protein,
CC described in the invention
XX
SQ Sequence 1299 BP; 317 A; 345 C; 355 G; 282 T; 0 U; 0 Other;
Query Match 17.0%; Score 224.6; DB 3; Length 1299;
Best Local Similarity 53.2%; Pred. No. 3.4e-62;
Matches 571; Conservative 0; Mismatches 489; Indels 13; Gaps 4;
Qy 107 TAGAACGTGTGTTATTTTGTAGTCCCATGTTGCTCGCCGACCAA-ACAACGACG 165
Db 89 TGGAAAGTGGTGATGTCAGTCGTGCTGCTGCTCCAAACCAAGCCACGCAAC 148
Qy 166 TTATGAATGATGTTACGCCAGATAGTGGCCACATAGTCCCGGTAAGAGCGGGTATTAA 225
Db 149 TGATGAGGATGTCACCCAGACGATGGCCACCTGGCCGTAAGAACTGGGTGGCTGA 208
Qy 226 CGCCACGTGGTGGGAGTTGTCATGATGGGGGGTTTTATGCTGATTAATTCTCGCA 285

Db 209 CACCGCGCGGTGGTGGAGTAATCGCCTATCTCGGACATTACCAACGCCAGCGTCTGGTAG 268
QY 286 GCGTTGGTTTGTAGCG---GCGGATGTCCGGCAGAGGGGGGTATATGACACAGGCGAG 342
Db 269 CCGACGGATTGCTGGCGAAAGAGGCTGCCCGCAGTCTGGTCAGGTCCGGATTATGCTG 328
QY 343 ATATCGATCAACGATACCGCTTAACCGGACAGGCAATTTCTTGATGGTGTGCTCGGGGT 402
Db 329 ATGTGACAGCGGTACCGCTTAACCGGAGGAGCCTTCGCCCGCGGGCTGGACCTGACT 388
QY 403 GTGGTTTGGACCGTCATATCAGCGCGATTTGAAAGAACCGATCCCTGTTCATCCGG 462
Db 389 GTGCAATACCGTACATACCCAGGCGAGATGCTCCGATCCGATCGTTATTTAATCTC 448
QY 463 TAGAGACTGGGTGTGTAGTATGACCAACGACCAACAGATTAACCGATTGAGNACGAT 522
Db 449 TAAAACTGGCGTTTGGCCAACTGGATAACGCGAAGCTGACGCGATCTCTCAGCAGG 508
QY 523 TGGCGGCGCGTTAGATACGGTGGAGCCAGCGCTAGCTTAAACCTTTGCCAGATGGGG 582
Db 509 CAGGAGGTGCAATGCTGACTTACCGGCGCATCGCAACAGCGCTTCGCGACTGGAC 568
QY 583 AGTGTGTAATTTGGCGGCTTCTCTTATGCAATCTTTGCAACAGCAAGCAAAACGT 642
Db 569 GGGTGTAAATTTCCGCAATCAAACTTGTGCTTAAACGCTGAGAAACAGGACGAAAGCT 628
QY 643 GTGATTTGCCCACTTTGGCGCAATGAAGTTAAAGTTAATAAGAGGACAAAGTGA 702
Db 629 GTTCATTAACGAGCATTAACATCGAATCAAGTGAAGCCGA-----CATGCT 682
QY 703 CCTCAGTGGCCACTGGCGTTATCATCGACATTTGGGTGTAATCTTTTATTAACAAACG 762
Db 683 CATTAACCGGTGGGTGAAGCTCGCATCAATCTGACGGAGATATTTCTCTGCAACAG 742
QY 763 CACAAGCCATGCCAGAGTTGCTGCAACGCTTAAAGGGCGGAGATTGGGTATCCT 822
Db 743 CACAGGATGCGGAGCGGGGTGGGAGGATCACCGATTACACCATGGACACT 802
QY 823 TATTGTCAATACAGCGCAATTTAATTAATGCGCAAAACACGATATGCGCCGTC 882
Db 803 TGCTAAGTTTGCATAACGGCAATTTATTTGCTACAGCGCAGGAGTTGCCCGCA 862
QY 883 ATAAAGGAGCGCATTAATACAGCAGATAGATACGCTTTAACCTTCACTGGATGTC 942
Db 863 GCGCGCCACCGCTTATTAGATTGATCAAGACAGCTTGACGCCCATC---CACGCG 919
QY 943 AGGGGCAAAAGCTACCCATTTCAGCCCAAAACCGGGTCTTTGCTCGGTGGGCGATGATA 1002
Db 920 AAAAAACAGGCGTATGGTGTGACATTACCCACTTCAGTGTCTTTATCGCGGACACGATA 979
QY 1003 CCAATATGCAATATGCGGTATCTGGAGCCGACTGGCAGCTACCGGAGCACTG 1062
Db 980 CTAATCTGGCAATCTCGCGCGGCACTGGAGCTCAACTGGAGCTTCCCGGTGACCGG 1039
QY 1063 ATAATACTCCACAGGTGGGGGATTTGGTTTTTGAACATATGGCAAAATCCGGATAACCC 1122
Db 1040 ATAAACGCGCGCAGGTGGTGAACCTGGTGTGAAACGCTGGCTGGCTAAGCATAACA 1099
QY 1123 AGCGCTACGTTGCGGTGAAGATGTTCTACCAACGATGATCGATGTCGTAAT 1175
Db 1100 GCAGTGGATTACAGTTTCGCTGCTTCCAGACTTTACAGCAGATGCGGTAT 1152

RESULT 13
ACA19297
ID ACA19297 standard; DNA; 1299 BP.
XX AC
XX ACA19297;
XX
DT 19-JUN-2003 (first entry)
XX
DE Prokaryotic essential gene #954.
XX

KW Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
OS Escherichia coli.
XX WO300277183-A2.
XX
XX 03-OCT-2002.
XX
XX 21-MAR-2002; 2002WO-US009107.
XX
XX 21-MAR-2001; 2001US-00815242.
XX 06-SEP-2001; 2001US-00948993.
XX 25-OCT-2001; 2001US-0342923P.
XX 08-FEB-2002; 2002US-00072851.
XX 06-MAR-2002; 2002US-0362899P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
XX WPI; 2003-029926/02.
XX P-PSDB; ABU15427.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
XX for homologous nucleic acids required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 14; SEQ ID NO 7167; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-required gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX compound's activity; (11) a culture comprising strains in which the gene
XX product is overexpressed or underexpressed; (12) determining the extent
XX to which each of the strains is present in a culture or collection of
XX strains; or (13) identifying the target of a compound that inhibits the
XX proliferation of an organism. The antisense nucleic acids are useful for
XX identifying proteins or screening for homologous nucleic acids required
XX for cellular proliferation to isolate candidate molecules for rational
XX drug discovery programs, or for screening homologous nucleic acids
XX required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
XX *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
XX prokaryotic essential genes. Note: The sequence data for this patent did
XX not form part of the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences

Query Match 17.0%; Score 224.6; DB 7; Length 1299;
Best Local Similarity 53.2%; Pred. No. 3.4e-62;
Matches 571; Conservative 0; Mismatches 489; Indels 13; Gaps 4;
QY 107 TAGAACGTGTGTTATTTTGGTCCCATGTTGCTCGCCACCAA-ACAACGAGC 165
DB 89 TGGAAAGTGTGTTGATTGTCATGTCATGTCGTGCTCCCAACCAAGCCACGCAAC 148

166	Qy	TTATGAATGATGTTATCGCCAGATTAAGTGGCCACAATGCGCGGTAAAAAGCGGGTATTATAA	225
167			
168			
169			
170			
171			
172			
173			
174			
175			
176			
177			
178			
179			
180			
181			
182			
183			
184			
185			
186			
187			
188			
189			
190			
191			
192			
193			
194			
195			
196			
197			
198			
199			
200			
201			
202			
203			
204			
205			
206			
207			
208			
209			
210			
211			
212			
213			
214			
215			
216			
217			
218			
219			
220			
221			
222			
223			
224			
225			
226			
227			
228			
229			
230			
231			
232			
233			
234			
235			
236			
237			
238			
239			
240			
241			
242			
243			
244			
245			
246			
247			
248			
249			
250			
251			
252			
253			
254			
255			
256			
257			
258			
259			
260			
261			
262			
263			
264			
265			
266			
267			
268			
269			
270			
271			
272			
273			
274			
275			
276			
277			
278			
279			
280			
281			
282			
283			
284			
285			
286			
287			
288			
289			
290			
291			
292			
293			
294			
295			
296			
297			
298			
299			
300			
301			
302			
303			
304			
305			

RESULT 14
AAC68298
ID AAC6
XX

AC	AAC68298;
XX	
XX	15-SEP-2003 (revised)
DT	20-FEB-2001 (first entry)
XX	
DE	R15/APPA plasmid coding sequence.
XX	
KW	Transgenic animal; salivary protein; phytase; phosphorus; animal growth;
KW	environmental pollution; pig; ds.
XX	
OS	Rattus sp.
OS	Escherichia coli.
OS	Chimeric.
XX	
FN	WO200064247-A1.
XX	
PD	02-NOV-2000.
XX	
PF	20-APR-2000; 2000WO-CA000430.
XX	
PR	23-APR-1999; 99US-0130508P.
XX	
PA	(UYGU-) UNIV GUELPH.
XX	
PI	Forsberg CW, Golovan S, Phillips JP;
XX	
DR	WPI; 2000-687245/67.
DR	P-PSDB; AAB36261.
XX	
PT	Transgenic non-human animal for gastrointestinal tract specific
PT	expression of a protein, preferably phytase, comprises a nucleic acid
PT	sequence including a heterologous transgene construct encoding the
PT	protein.
XX	
PS	Claim 14; Fig 21; 152pp; English.
XX	
CC	The present invention provides transgenic animals which produce desired
CC	proteins, in this case pigs which expresses phytase in the salivary
CC	Gland. Low phytase production levels result in phytate in the diet being
CC	excreted and causing phosphorus contamination in water, as well as
CC	reducing the growth of animals. The invention provides a number of
CC	transgenics containing the E. coli APPA phytase coding sequence. (Updated
CC	on 15-SEP-2003 to standardise OS field)
XX	
SQ	Sequence 3470 BP; 1065 A; 721 C; 735 G; 949 T; 0 U; 0 Other;
	Query Match 17.0%; Score 224.6; DB 3; Length 3470;
	Best Local Similarity 53.2%; Pred. NO. 6.1e-82;
	Matches 571; Conservative 0; Mismatches 489; Indels 13; Gaps 4;
Qy	107 TAGAACGTGCTGGTTATTTTGAAGTCGCATGGTGCTCGCTCGCGACCACAAACAACGCAGC 165
Dd	1899 TGCAAAGTGTGGTAGTTGTCACTGCTCATGGTGCTGCCAACCAAGGCCACGCAAC 1958
Qy	166 TTATGATGATGTTAGCCAGATAGTGGCCACATAGGCGGTAAAAGCGGGATTATAA 225
Dd	1959 TGTATGAGGATGTCACCCAGACGCATGCCCACCTGGCGCGTAAAAACTGGGTGGCTGA 2018
Qy	226 CGCACGTGTCGGGAGTGGTCACATTGATGGGGGGTTTTATGTGTGATTACTTTTCGCA 285
Dd	2019 CACCGCGGGTGTGAGCTAAATGCCCTATCTCGACATTACCAACGCCAGGCTCGTAG 2078
Qy	286 GCCTTGTTGTTAGCG---GGGGATGTCGCGCAGAGGGGGGGTATATGCAACGCAG 342
Dd	2079 CCACAGGATGCTGGCGGAAAAAGGGCTGGCCCGCAGTCTGCTCAGGTCGCGATTATTGCTG 2138
Qy	343 ATATCGATCAACTGACCCGCTTAACCGGACAGGCATTTCTTGATGTGTGGCTCCGGGGT 402
Dd	2139 ATGTGACGAGCGTACCCGTAAACAGCGAGCCTTGGCGCGGGCTGGCACTTGACT 2198
Qy	403 GTGGTTTGACCGTGCATAATCAGGCCGATTTGAAAAAGACCGATCCCGCTGTTCCATCCGG 462
Dd	2199 GTGCAATAACCGTACATACCCAGGACAGTACGTCCTAGTCCCGATCGGTATTATTAATCTC 2258

463	QY	TAGAGACTGGCGTGTGTAGTATGACACACGACAAACAGATAAAGCGATTGAAGACGAT	522
2259	Db	TAAAAACTGGCGTTTCCCAACTGATAACGCGAACGTGACTGACGCGATCTCTACGAGG	2318
523	QY	TGGCGGGCGGCTTAGATACGGTGAAGCAGCGCTACGCTAAACCTTTTGGCCAGATGGGG	582
2319	Db	CAGGAGGGTCAATTGCTGACTTTACCGGCATCGGCAACGGCGTTTCGGAACGGAAC	2378
583	QY	ACGTGCTGAATTTGGCGCTTCTCCTATTGCAATCTTTGCAACAGCAAGNAACCT	642
2379	Db	GGGTGCTTAATTTTTTCGCAATCAAACTGTGCTCTTAAACGTGAGAACAAGGACGAAGCT	2438
643	QY	GTGATTTGCGCACTTTGGCGCAATGAAGTTAACTTAATAAGAAAGGACAAAAGTGA	702
2439	Db	GTTCATTAACGCAAGGCATTACCATCGGAACCTCAAGGTGAGCGCCGA-----CAATGTCT	2492
703	QY	CCCTCAGTGGCCACTGGCGTTATCATCGACATTTGGTGAATCTTCTTATTACAAAAG	762
2493	Db	CATTAAACGGTGCAGTAAGCCTCGATCAATGCTGACGGAGATATTTCCTCTCAACAAG	2552
763	QY	CACAAGCCATGCCAGAGTTGCCTGGCAACGGCTAAAGGGCGGAGAAATGGGTATCCT	822
2553	Db	CACAGGAATGCGGAGCGCGGTGGGAAGGATCACCGATTCAACCAGTGGAAACACT	2612
823	QY	TATTGTCATTACATAAGCGCAATTAATTTAATGCAAAAACACCGTATATCGCCGTC	882
2613	Db	TGCTAAGTTGCAATACGCGCAATTTATTGCTCAACGACAGAGTTGCCGCA	2672
883	QY	ATAAAGGAGCCGCAATTATTACAGCAGATAGATACGGCTTTAACCTTCACTGGATGCTC	942
2673	Db	GC CGCGCCACCCGGTTATTAGATTGATCAAGACAGCGTTGACGCCCATC---CACCGC	2729
943	QY	AGGGGCAAAAGCTACCCATTTGACGCCAAAACCGGTTCTTGTCCTCGGTGGGCATGATA	1002
2730	Db	AAAAACAGGCGTATGGTGACATTACCACTTCAGTGTGTTTATGCGCGACACGATA	2789
1003	QY	CCAATATTGCCAATATTGCGGGTATGCTGGGAGCCGACCTGGCAGCTACCGAGCAACCTG	1062
2790	Db	CTAATCTGGCAAAATCTCGCGGGCGCATGGAGGCTCACTGGACGCTTCCGGTCAAGCGG	2849
1063	QY	ATAATACTCCACAGGTGGGGGATGGTTTTTGAATATGCGCAAAATCCGGATAACCAAC	1122
2850	Db	ATAACAGCGCCAGTGGTGAATGGTGGTTTGNACCTGGCGCTCGGCTAAGCGATAACA	2909
1123	QY	AGGCTACGTTGCGGTGAAGATGTTCACAAACGATGGATCGATGTCGTGAT	1175
2910	Db	GCCAGTGGATTCAGGTTTCGCTGGCTTCAGACTTTACAGCAGATGCGTGAT	2962

RESULT 15
AAC68296
ID AAC68296 standard; DNA; 4060 BP.

XX	
AC	AAC68296;
XX	
DT	15-SEP-2003 (revised)
DT	20-FEB-2001 (first entry)

XX R15/APPA plasmid coding sequence.

Transgenic animal; salivary protein; phytase; phosphorus; animal growth; KW
 environmental pollution; pig; ds. KW

OS	Rattus sp.
OS	Escherichia coli.
OS	Chimeric.

XX PN WO200064247-A1.

XX
PD 02-NOV-2000.

AA
PF 20-APR-2000; 2000WO-CA000430.

XX	23-APR-1999;	99US-0130508P.	
XX	(UYGU-) UNIV GUELPH.		
XX	Forsberg CW, Golovan S, Phillips JP;		
XX	WPI: 2000-687245/67.		
XX	P-PSDB; AAB36259.		
XX	Transgenic non-human animal for gastrointestinal tract specific		
XX	expression of a protein, preferably phytase, comprises a nucleic acid		
XX	sequence including a heterologous transgene construct encoding the		
XX	protein.		
XX	Claim 14; Fig 19; 152pp; English.		
XX	The present invention provides transgenic animals which produce desired		
XX	proteins, in this case pigs which expresses phytase in the salivary		
XX	gland. Low phytase production levels result in phytate in the diet being		
XX	excreted and causing phosphorus contamination in water, as well as		
XX	reducing the growth of animals. The invention provides a number of		
XX	transgenes containing the E. coli APPA phytase coding sequence. (Updated		
XX	on 15-SEP-2003 to standardise OS field)		
XX	Sequence 4060 BP; 1257 A; 814 C; 843 G; 1146 T; 0 U; 0 Other;		
XX	Query Match 17.0%; Score 224.6; DB 3; Length 4060;		
XX	Best Local Similarity 53.2%; Pred. No. 6.6e-62;		
XX	Matches 571; Conservative 0; Mismatches 489; Indels 13; Gaps 4;		
QY	107 TAGAACGTGTGGTTATTTGAGTCGCGCATGGTTCGCTCGCGCACCA-ACAAACGCAGC 165		
DB	1899 TGAAGAAGTGGTGAATGTCAGTCGTCATGGTGTGCGTGTCTCCACCAAGCCACGCAC 1358		
QY	166 TTATGAATGATGTTACGCCAGATAAAGTGGCCACAATGGCGGTTAAAAGCGGGGTATTTAA 225		
DB	1959 TGATGCAGGATGTCACCCAGACGCATGGCCACCTGGCCGGTAAAACCTGGGTGGCTGA 2018		
QY	226 CGCACGTGGTGGGAGTTGGTCACATTGATGGGGGGTATTATGTGATTACTTTCGCA 285		
DB	2019 CACCGCGCGTGGTGAAGTAATCGCCTATCTCGGACATTACCAACGCCAGCTCTGGTAG 2078		
QY	286 GCCTTGGTTGTTAGCG---GCGGGATGTCGGGCAGAGGGGGGGGTATATGCACAGGCAG 342		
DB	2079 CCAGCGGATGCTGGCGGAAAAGGGCTGCCCGCAGTCTGTCAGGTCGGATTATTCGTG 2138		
QY	343 ATATCGATCAACGTACCCGCTTAACCGGACAGGCATTTCTTGATGTGTGTGGTCCGGGGT 402		
DB	2139 ATGTCAGCAGCGTACCCCGTAAACAGCGGAACCCCTTCGCGCGGGGCTGGCACCTGACT 2198		
QY	403 GTGGTTTGACCGTGCAATAATCAGCGCCGATTTGAAAAGACCGATCCCTCTTCATCCGG 462		
DB	2199 GTGCAATAACCGTACATACCCAGCGAGATAGTCTCCAGTCCCGATCGGTATTTAATCCTC 2258		
QY	463 TAGAGACTGCGGTGTGTAGTTAGACAACGCACAACAGATAAAGCGATTGAAGAACGAT 522		
DB	2259 TAAAAACTGCGGTTTGCCAACTGATAACCGCAACGTGACTGACGCGATCCTCAGCAGGG 2318		
QY	523 TGGCGCGCGGTAGATACGGTCAGCGCGGTACGCTAAACCTTTTGGCCAGATGGGGG 582		
DB	2319 CAGAGGGTCAATTGCTGACTTTACGGGGCATCGGCAACGGGTTTCGGACTGGAAC 2378		
QY	583 ACGTGCTGAATTTTGGCGGCTTCTCTTATTGCAAAATCTTTTGCAACAGCAAGGAAAAACGT 642		
DB	2379 GGGTGTCTTAATTTTCCGAATCAAACTGTGTCCCTTAAACCGTGAGAAACAGGACGAAAGCT 2438		
QY	643 GTGATTTTGGCCACTTTGGCGCCAAATCAAGTTAACGTTTAATAAGAGGGGACAAAAGTGA 702		
DB	2439 GTTCATTAAACGCGAGGCATTAACCATCGGAATCTCAAGGTGACGGCGGA-----CAATGTCT 2492		
QY	703 CCCTCAGTGGGCGCACTGGGCTTATCATCGCAATTTGGGTGAAATCTTCTTATTACAAAACG 762		

Db	2493	CATTAAACCGGTGCGGTAAAGCCTGCAATCAATGCTGACGGAGATATTTCTCTGCAACAAG	2552
Qy	763	CACAAGCCATGCCAGAGGTTGCCCTGGCAACGGCTAAAAGGGGGGAGATTGGGTATCCT	822
Db	2553	CACAGGGNATGCCGGAGCCGGGTGGGGAAGGATCACCGATTACACCAAGTGGACACCT	2612
Qy	823	TATTGTCATTACATAACGGCAATTTAATTTAATGGCAAAAACACCGTATATCGCCGTC	882
Db	2613	TGCTAAGTTTGCATAACCGCAATTTTATTGTCTACAAACGACGCCAGAGGTTGCCGCA	2672
Qy	883	ATAAAGGACGCCATTATTACAGCAGATAGATACGGCTTTAAACCTTCAACTGGATGCTC	942
Db	2673	GCCGCGCCACCCGTTATTAGATTGATCAAGACAGCGTTGACGCCCCATC---CACCGC	2729
Qy	943	AGGGGCAAAAGCTACCCATTTACGCCCAAAACCGGGTCTTGTTCCTCGTGGGATGATA	1002
Db	2730	AAAAACAGCGCGTATGGTGTGACATTACCCACTTCAGTGTCTTTTATCGCCGGACACGATA	2789
Qy	1003	CCAAATTTGCCAATATTGCGGGTATGTTGGAGCCGACTGCGAGCTACCCGAGCAACCTG	1062
Db	2790	CTAATCTGGCAAAATCTCGCGCGCGACTGGAGCTCAACTGAGCCCTTCCCGGTACGCCG	2849
Qy	1063	ATAATCTCCACCGAGTGGGGATTGGTTTTTGAACATAATGGCAAAATCCGGATAACCAAC	1122
Db	2850	ATAACAGCCGCCAGGTGGTGAACCTGGTGTGTTGAACGCTGGCGTCCGCTAAGCGATAACA	2909
Qy	1123	AGCGCTAGCTTGGCGTGAAGATGTTCTACCAACGATGGATCAGTTGGGTAAT	1175
Db	2910	GCCAGTGGATTACAGTTTCGCTGGTCTTCCAGACTTTACAGCAGATGCGTGAT	2962

Search completed: May 2, 2004, 22:50:08
 Job time : 382.238 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.
OM nucleic - nucleic search, using sw model
Run on: May 2, 2004, 22:22:03 ; Search time 72.6484 Seconds
(without alignments)
10121.490 Million cell updates/sec
Title: US-10-021-723B-3
Perfect score: 1325
Sequence: 1 atgtcagattagaaatcg.....aacctgctgcacatctaa 1325
Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0
Searched: 682709 seqs, 277475446 residues
Total number of hits satisfying chosen parameters: 1365418
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : Issued Patents NA:*
1: /cgn2_6/prodata/2/ina/5A-COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B-COMB.seq:*
3: /cgn2_6/prodata/2/ina/6A-COMB.seq:*
4: /cgn2_6/prodata/2/ina/6B-COMB.seq:*
5: /cgn2_6/prodata/2/ina/6C-COMB.seq:*
6: /cgn2_6/prodata/2/ina/6D-COMB.seq:*
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

ALIGNMENTS

RESULT 1
US-09-259-214-1
; Sequence 1, Application US/09259214A
; Patent No. 6110719
; GENERAL INFORMATION:
; APPLICANT: Kretz, Keith
; TITLE OF INVENTION: NOVEL PHYTASE
; FILE REFERENCE: DIVER1370-1
; CURRENT APPLICATION NUMBER: US/09/259,214A
; EARLIER FILING DATE: 1999-03-01
; EARLIER APPLICATION NUMBER: 08/910,798
; EARLIER FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1323
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1320)
; NAME/KEY: misc feature
; LOCATION: (1)...(1323)
; OTHER INFORMATION: n = A,T,C or G
US-09-259-214-1

Query Match 16.5%; Score 218.8; DB 3; Length 1323;
Best Local Similarity 52.8%; Pred. No. 1.4e-63;
Matches 567; Conservative 0; Mismatches 493; Indels 13; Gaps 4;
QY 107 TAGAACGTGTGTTTATTTAGTCCATGCTGCTCGCCGACCAA-ACAAACGAGC 165
DB 89 TGGAAAGTGTGTTGATTTGTCAGTCTGCTGCTCCACCAAGCCACGCAAC 148
QY 166 TTATGATGATGTACGCCAGATAAGTGGCCACAAATGCGCGGTAAAGCGGGGTATTAA 225
DB 149 TGATGCGAGATGTACCCGACGACATGCGCCGCTGAACTGGGTGGGTGA 208
QY 226 CGCCACGTGTGCGAGTTGTTGATGGGGGGGTTTATGTTGATTTACTTTGCA 285
DB 209 CACCGCGGTGTGTTGAGTATTCGCTATCTCGACATTACCAACGCGGCTGTTAG 268
QY 286 CCCTTGGTTTGTAGCG---CGGGATGTCCGCGAGAGGGGGGGGTATATCGACAGGAG 342
DB 269 CCGACGATTTCTGCGGAAAAGGGCTGCCCGAGTCTGGTTCAGGTCCGATTATTGCTG 328
QY 343 ATATCGATCAGGTACCGCTTAACCGACAGGAGGATTTCTTGATGGTGTGCTCGGGGT 402
DB 329 ATGTCGACGCGTACCGTAAACAGCGAGCCCTTCGCCCGCGGCTGGCACTGACT 388

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	218.8	16.5	1323	3	US-09-259-214-1
2	218.8	16.5	1323	3	US-09-318-528-1
3	218.8	16.5	1323	3	US-09-291-931-1
4	218.2	16.5	1489	4	US-09-540-149A-9
5	143.6	10.8	1272	2	US-08-910-798-1
6	53.8	4.1	1266	4	US-09-489-039A-341
7	48	3.6	1569	4	US-09-489-039A-6330
8	48	3.6	5975	1	US-08-920-812-23
9	48	3.6	5975	1	US-08-920-827-23
10	48	3.6	5975	1	US-08-177-23
11	48	3.6	5975	1	US-08-362-577C-23
12	48	3.6	5975	2	US-08-920-828-23
13	37	2.8	7218	1	US-08-232-463-14
14	35	2.6	1602	4	US-09-4062-4062
15	35	2.6	1773	4	US-09-252-991A-4093
16	34.2	2.6	402	4	US-08-821-976-17355
17	34.2	2.6	5185	4	US-09-596-594-640
18	33.8	2.6	9775	4	US-09-647-540A-1
19	33.8	2.6	9775	4	US-09-647-540A-3
20	33.8	2.6	9775	4	US-09-647-540A-5
21	33.8	2.6	9775	4	US-09-647-540A-7
22	33.8	2.6	9775	4	US-10-119-600-1
23	33.8	2.6	9775	4	US-10-119-600-3
24	33.8	2.6	9775	4	US-10-119-600-5
25	33.8	2.6	9775	4	US-10-119-600-7
26	33.8	2.6	9775	4	US-10-119-651-1
27	33.8	2.6	9775	4	US-10-119-651-3

```
QY 403 GTGGTTACCGTCATTAATCAGCGCGATTGAAAGACCGATCCCTGTTCCATCCGG 462
Db 389 GTGCATTAACCGTACATACCAGCGAGATAGCTCAGTCCCGATCCGTTATTAATCCTC 448
QY 463 TAGAGACTGGCGTGTGAAGTTAGACAAACGACAAACAGATAAAGCGGATGAAGACGAT 522
Db 449 TAAAACTGGCGTTTGCCAACTGGATAACGCGAAGCTGACTGACGCGATCCTCAGCAGGG 508
QY 523 TGGCGGCGCGTTAGATACGGTGAGCCAGCGCTAGCTAGCTAAACCTTTGCCCGAGATGGGG 582
Db 509 CAGGAGGTCAATTGCTGACTTTACCGGCAATCGGCAACCGGCTTTGCGAACTGGAAC 568
QY 583 ACGTCTGAATTTTGGCGCTTCTCCTTATTCGCAATCTTTTCCAAACAGCAAGGAAACAGT 642
Db 569 GGGTCTTAAATTTTCCGCAATCAAACTTGCTCTTAAACGCTGAGAAACAGGAAAGCT 628
QY 643 GTGATTTTGGCCACTTTCCGCGCAATGAAGTTAAGTTAATTAAGAGAGGACAAAGTGA 702
Db 629 GTTCAATTAACGAGGCAATACCATCGGAACCTCAAGGTGAGCGCGA-----CAATGTCT 682
QY 703 CCTCAGTGGGCACTGGCGTTATCATCGACATTTGGGTGAATCTTCTTATTAACAAACG 762
Db 683 CATTAACCGGTGCGGTAGCGCTCGCATCAATGCTGACGAGATATTTCTCTGCAACAG 742
QY 763 CAAAGCCATGCGAGGTTGCTTGCGCAACGCTTAAAGGGGCGGAGAAATGGGTATCCT 822
Db 743 CACAGGGAATGCGGAGCGGGGTGGGAAGGATCACCGATTACACCGTGAACACCT 802
QY 823 TATTGTCAATACATAACCGCAATTTAATTTAATGCGCAAAACACCGGTATATCCCGGTC 882
Db 803 TGTCAAGTTTGCATTAACCGGCAATTTTATTTGCTACACGCAAGCGAGAGTTGCCGCGCA 862
QY 883 ATAAAGGACGCGCATTAATACGACAGATAGATAGCGCTTTAAACCTTCAACTGAGATGCTC 942
Db 863 GCCGCGCCACCCCGTTATTGGATTGATCATGGCAGCGTTGACGCGCCATCCAC---CGC 919
QY 943 AGGGCAAAAGCTACCCATTTCAGCCAAACCGGGTCTTCTCTCGGTGGCATGATA 1002
Db 920 AAAAAACGCGTATGGTGTGACATTAACCCATTTCAGTACTGTTTATTTCCGAGACAGATA 979
QY 1003 CCAATATTGCCAATATTCGGGTATGCTGGAGCGGACTGGCAGCTACCGGAGCAACCTG 1062
Db 980 CTAATCTGCAAACTCTCGCGCGCACTGGAGCTCAACTGGACGCTTCCCGGTGAGCGGG 1039
QY 1063 ATAATACCTCCACGAGTGGGATGTTTGTGAACTATGCAAAATCCGATACCAACC 1122
Db 1040 ATAAACGCGCGCAGGTGTGAACTGTTTGAACGCTGAGTGTGAACGCTGGGCTAAGCGATAACA 1099
QY 1123 AGCGTACGTTGCGGTGAAGATGTTCTACCAACGATGGATCAGTTGCGTAAT 1175
Db 1100 GCCAGTGGATTGAGTTTGGCTGCTTCCAGACTTTACAGCAGATGCGTGTAT 1152
```

RESULT 2

```
US-09-318-528-1
; Sequence 1, Application US/09318528
; Patent No. 6183740
; GENERAL INFORMATION:
; APPLICANT: Kretz, Keith
; TITLE OF INVENTION: NOVEL PHYTASE
; FILE REFERENCE: 09010/029003
; CURRENT APPLICATION NUMBER: US/09/318,528
; CURRENT FILING DATE: 1999-05-25
; EARLIER APPLICATION NUMBER: 09/291,931
; EARLIER FILING DATE: 1999-04-13
; EARLIER APPLICATION NUMBER: 08/910,798
; EARLIER FILING DATE: 1997-08-13
; EARLIER APPLICATION NUMBER: 09/259,214
; EARLIER FILING DATE: 1999-03-01
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1323
```

```
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1320)
; NAME/KEY: misc feature
; LOCATION: (1)...(1323)
; OTHER INFORMATION: n = A,T,C or G
US-09-318-528-1
```

```
Query Match 16.5%; Score 218.8; DB 3; Length 1323;
Best Local Similarity 52.8%; Pred. No. 1.4e-63;
Matches 567; Conservative 0; Mismatches 493; Indels 13; Gaps 4;
```

```
QY 107 TAGAAGCTGTGTTATTTTGGTTCGCAATGTTGCTCGCGCACCAA-ACAACGCGACG 165
Db 89 TCGAAGTGTGTTGATTTGTCAGTGTGCTGCTCCACCAAGGCGACGCAAC 148
QY 166 TTATGAATGATTTAGCGCCAGATAAGTGGCCCAATATGCGCGGTAAAGCGGGGTATTTAA 225
Db 149 TGTATGAGGATGTCAACCCAGACGATGCGCAACCTGCGCGGTAAACCTGGGTGCTGA 208
QY 226 CGCCAGCTGTGCGGAGTTGTCACATTTGATGGGGGGTTTTATGTTGATTTACTTTCCGA 285
Db 209 CACCGGNGGTGTGAGCTAATCGCTATCTCGACATTACCAAGCCAGCGTCTGGTAG 268
QY 286 GCCTTGGTTTGTAGCG---GCGGATGTCGCGCAGAGGGGGGTATATGCACAGGCGAG 342
Db 269 CGGACGATTGTGCGGCAAAAGGGTGCCTGCTGTCAGTCTGTCAGGCTCGCGATTATTGCTG 328
QY 343 ATATCGATCAACGTACCGCTTAAACCGACAGGCAATTTCTGATGTTGGTCCGGGT 402
Db 329 ATGTCACAGGCGTACCGCTTAAACAGGCAAGGCTTTCGCCCGGGGTGGCAGCTGACT 388
QY 403 GTGTTTGGCCGTGATATCAGGCGGATTTGAAAGAACCGATCCCTGTGTTCCATCCGG 462
Db 389 GTGCAATAACCGTACATACCCAGGCGATACGTCAGTCCGATCCGTTATTTAATCTCTC 448
QY 463 TAGAGACTGGCGTGTGAAGTTAGACACGCAACAGATAAAGCGTTGAAGACGAT 522
Db 449 TAAAACTGGCGTTTGCCAACTGGATACGCGAAGCTGACTGACGCGATCCTCAGCAGGG 508
QY 523 TGGCGGCGCGTTAGATACGCTGAGCAGCGCTAGCTTAAACCTTTTGCAGATGGGGG 582
Db 509 CAGGAGGTCAATTGCTGACTTTACGGGCATCGGCAACCGGTTTCGCAACTGGAAC 568
QY 583 ACGTCTGAATTTTGGCGGTCTCCTTATGCAATCTTTGCAACAGCAGGAGAAAACGT 642
Db 569 GGGTGTCTTAATTTTCCGCAATCAAACTTGTGCTTTAAACGTCGAGAAACAGGACGAAAGCT 628
QY 643 GTGATTTTGCCTACTTTGCGGCCAATGAAGTTAAACGTTAATAAGAGAGGACAAAAGTGA 702
Db 629 GTTCATTAACGAGGCAATTACCATCGGAATCAGGTGAGCGCGA-----CAATGTCT 682
QY 703 CCTCAGTGGGCGACTGGCGTTATCATGACATTTGGGTGAATCTTCTTATTAACAAACG 762
Db 683 CATTAACCGGTGCGGTGAGCGGCTGGGAGAGATCACCGATTACACCGAGTGAACACCT 742
QY 763 CACAAGCCATGCGCAGAGTTGCTGCGCAACGCTTAAAGGGGCGGAGAAATTTGGGTATCCT 822
Db 743 CACAGGGAATGCGGAGCGGGGTGGGAGAGATCACCGATTACACCGAGTGAACACCT 802
QY 823 TATTGTCAATTAACAGCGCAATTTAATTTAATGCGCAAAACACCGTATATCGCGGTGTC 882
Db 803 TGCTAAGTTTGCATAACGCGCAATTTTATTTGCTACACGCAAGCGAGAGTTGCCGCGCA 862
QY 883 ATAAAGGAGCGCATTTATACAGCAGATAGATAGCGCTTTAAACCTTCAACTGAGTGTCTC 942
Db 863 GCGCGCGCCACCCCGTTATTGGATTGATCATGGCAGGTTGAGCGCCATCCAC---CGC 919
QY 943 AGGGGCAAAAGCTACCCATTTCAGCCCAAAACCGGGTCTTGTCTCTCGGTGGCATGATA 1002
Db 920 AAAAAACGCGTATGGTGTGACATTAACCCATTTCAGTACTGTTTATTTCCGAGACAGATA 979
```

QY 1003 CCAATATTCCTATTTGGGATGCTGGAGCCGACTGCTACCCGAGCACTG 1062
DB 980 CTAATCTGGCAATCTCGCGCGGCACTGAGCTCACTGGACGCTTCCCGTCAACCGG 1039
QY 1063 ATAATACTCCACAGTGGGGATTTGGTTTGAATATGCAAAATCCGATAACCAACC 1122
DB 1040 ATAACACGCGCGAGTGGTGAATGTTTGAACGCTGGCGTCCGCTAAGCGATAACA 1099
QY 1123 AGCGTACCTTGGGCTGAAGATGTTCTACCAACGATGATCAGTTGGTAAAT 1175
DB 1100 GCCAGTGGATTCAGGTTTCGCTGGTCTCCAGACTTTACAGCAGATGGGTGAT 1152

RESULT 3
US-09-291-931-1
; Sequence 1, Application US/09291931A
; Patent No. 6190897
; GENERAL INFORMATION:
; APPLICANT: Kretz, Keith
; TITLE OF INVENTION: NOVEL PHYTASE
; FILE REFERENCE: 09010/029003
; CURRENT APPLICATION NUMBER: US/09/291,931A
; CURRENT FILING DATE: 1999-04-13
; EARLIER APPLICATION NUMBER: 08/910,798
; EARLIER FILING DATE: 1997-08-13
; EARLIER APPLICATION NUMBER: 09/259,214
; EARLIER FILING DATE: 1999-03-01
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1323
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1320)
; NAME/KEY: misc_feature
; LOCATION: (1)...(1323)
; OTHER INFORMATION: n = A,T,C or G
US-09-291-931-1

Query Match 16.5%; Score 218.8; DB 3; Length 1323;
Best Local Similarity 52.8%; Pred. No. 1.4e-63;
Matches 567; Conservative 0; Mismatches 493; Indels 13; Gaps 4;

QY 107 TAGAAGTGTGGTATTTAGTGGCCATGGTTCGCTCGCGACCAA-ACAAGCGAGC 165
DB 89 TGGAAAGTGTGGTATTTAGTGGCCATGGTTCGCTCGCGACCAA-ACAAGCGAGC 148
QY 166 TTATGATGATGTTAGCCGAGTAACTGGCCCAATGGCCGCTTAAAGCGGGTATTAA 225
DB 149 TGATGAGGATGTCACCCAGAGCGATGGCCAACTGGCCGCTTAAAGCGGGTATTAA 208
QY 226 CGCCAGTGTGGGAGTTGGTCAATGATGGGGGGTTTATGGTGTATTACTTTGCGCA 285
DB 209 CACCGGNGGTGGTGGCTAATCGCTATCTCGGACATTACCAACCCAGCGTCTGGTAG 268
QY 286 GCCTTGGTTTGTAGG---GCGGATGTCGCGACAGAGGGGGGTATATGACAGGCGAG 342
DB 269 CCGACGGATGCTGGCGGAAAGAGGGTGGCCCGAGTCTGTGTCAGGTGGCGATTTGCTG 328
QY 343 ATATCGATCAACGTAACCGGTTAAACCGGACAGGCAATTTCTTGATGTGTGGTTCGCGGT 402
DB 329 ATGTCAGAGCGTACCCGTTAAACAGGCGAGGCTTCGCGCGCGGCTGGCACCTGACT 388
QY 403 GTGGTTGACCGTGCAATATCAGCGCGATTTGAAAGACCGATCCCTGTTCCATCCGG 462
DB 389 GTGCAATAACCGTACATACCCAGCGAGATACGTCACGTCGCCGATCGTTATTATTCCTC 448
QY 463 TAGAGACTGGCGTGTGTAGTTAGCAACGCAACAGATAAAGCGATTTGAAGACGAT 522
DB 449 TAAAACTGGCGTTTCCCACTGGATACCGGAACGTCGCTGACGCGATCCTCAGCAGGG 508

QY 523 TGGGCGGCGCTTAGATACGGTGGAGCCGCTACGCTAAACCTTTTCCAGATGGGG 582
DB 509 CAGAGGGCTCAATGCTGACTTTACCGGGCATGGGAAACGGGCTTTCGGAAC 568
QY 583 ACGTGCTGAATTTTGGCGCTTCTCTTATTGCAAAATCTTTGCAACAGCAAGAAAAAGT 642
DB 569 GGGTGTCTTAATTTTCCGCAATCAAACTTGTGCTTAAACGTTAGAAACAGGACGAAAGCT 628
QY 643 GTGATTTTGGCCACTTTGGCGGCAATGAAGTTAAGTTAATAAAGAGGACAAAGTGA 702
DB 629 GTTCATTAAACGACGGCATTTACCATCGGAATCAAGGTGAGCGCGA-----CAATGCT 682
QY 703 CCCTCAGTGGGCCACTGGCGTTATCATCGACATTTGGGTGAAATCTTCTTATTCAAAACG 762
DB 683 CATTAAACGGTGGGTAAAGCCTCGCATCAATGCTGACGGAGATATTTCTCTGCAACAG 742
QY 763 CACAGCCATGCCAGAGTTGCTTGGCAACGGTAAAGGGGGGAGAAATGGGTATCTCT 822
DB 743 CACAGGGAATGCCGAGCGCGGGTGGGGAAGGATCACCGAATCACACAGTGAACACCT 802
QY 823 TATTGTCATTACATAACGCGCAATTTAATTTAATGCAAAACACCGTATATCGCCGTC 882
DB 803 TGCTAAGTTTGCATTAACGCGCAATTTTATTGCTACAAACGCGCAGAGGTTGCCGCA 862
QY 883 ATAAAGGAGCGCAATTTATACAGCAGATAGATAGGCTTTAAACCTTCAACTGATGCTC 942
DB 863 GCGCGCCACCGCTTATTGGATTTGATCATGGCAGCGTTGACGCGCCCATCCAC---CGC 919
QY 943 AGGGGCAAAAGACTACCCATTTAGCCCAAAACCGGCTTGTCTCGGTGGCGCATGATA 1002
DB 920 AAAACAGGGGTATGGTGTGACATTACCCACTTCACTGTTTATGCGGACAGATA 979
QY 1003 CCAATATTGCAATATTGCGGGTATGCTGGAGCCGACTGGCAGTACCCGAGCAACCTG 1062
DB 980 CTAATCTGGCAAAATCTCGCGCGGCACTGGAGTCACTGGAGCTTCCCGGTGAGCGG 1039
QY 1063 ATATACTCCACAGGTGGGGATTTGTTTGAATATGCAAAATTCGCGATAACCAAC 1122
DB 1040 ATAAACGCGCCAGGTGGTGAATGTTGAACTGTTGAACTGCTGGCTGCGCTAAGCGATA 1099
QY 1123 AGCCTACGTTGGCGTGAAGATGTTTACCAAAACGATGATCATGTTGCGTAAAT 1175
DB 1100 GCCAGTGGATTCAGGTTTCGCTGCTCTCCAGACTTTACAGCAGATGCGTAT 1152

RESULT 4
US-09-540-149A-9
; Sequence 9, Application US/09540149A
; Patent No. 6511699
; GENERAL INFORMATION:
; APPLICANT: Lei, Xingen
; TITLE OF INVENTION: ENZYMES WITH IMPROVED PHYTASE ACTIVITY
; FILE REFERENCE: 19603/2791
; CURRENT APPLICATION NUMBER: US/09/540,149A
; CURRENT FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 60/127,032
; PRIOR FILING DATE: 1999-03-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 9
; LENGTH: 1489
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-540-149A-9

Query Match 16.5%; Score 218.2; DB 4; Length 1489;
Best Local Similarity 52.8%; Pred. No. 2.4e-63;
Matches 567; Conservative 0; Mismatches 493; Indels 13; Gaps 4;
QY 107 TAGAAGTGTGGTATTTAGTGGCCATGGTTCGCTCGCGACCAA-ACAAGCGAGC 165
DB 270 TGGAAAGTGTGGTATTTAGTGGCCATGGTTCGCTCGCGACCAA-ACAAGCGAGC 329

QY 166 TTATGAATGATGTTACGCCAGATAGTGGCCACATAGTCCCGTAAAGCGGGGTATTAA 225
DB 330 TGATGAGATGTACCCAGACGATGCGCAACGTCGCCGCTAAACTGGGTTGGCTGA 389
QY 226 CGCCACGTGTCGCGAGTTGTCACATGATGGGGGGTATTTATGGTGATTAATTTCGCA 285
DB 390 CACCAACGCGTGGTGAGCTATCGCTATCTCGACATTAACAAGCCAGCGTCTGGTGG 449
QY 286 GCTTGGTTGTTAGC--GCGGATGTCGGCAGAGGGGGGGTATATCCACAGGAG 342
DB 450 CCGACGGATGTGCGCAAAAGGGCTGCCGACGCTGTGTCAGGTGCGGATTAATTGCTG 509
QY 343 ATATCGATCAACGTACCGCTTAACCGACAGGATTTCTTGATGGTGTGGCTCGGGGT 402
DB 510 ATGTCGACGAGTACCGTAAACAGAGCGAAGCTTCGCCGCGGGCTGCACTGACT 569
QY 403 GTGGTTGACGCGTGAATATCAGGCGGATTTGAAAGACGATCCCTGTTTCCATCCGG 462
DB 570 GTGCAATAACGTCATACCCAGGAGATGTCGAGTCCGATCCGTTATTTAATCCTC 629
QY 463 TAGAGACTGGGTGTTAGTTAGTAAAGCAGCAGCAACAGATTAAGCGGATTAAGAACGAT 522
DB 630 TAAAACTGGGTTTCCACTGCTGATACGGGACGTCAGTACGCGATCTCAGAGGG 689
QY 523 TGGGCGGCGGTAGATACGCTGAGCAGCGCTACGCTAAACCTTTTGGCCAGATGGGG 582
DB 690 CAGGAGGTCATCTGCTGACITTAACCGGCTACGCAAAAGCGGCTTCGCAACTGGAAC 749
QY 583 ACGTCTGAATTTGCGGCTCTCTTATTCGAATCTTTGCAACAGCAAGAAACGT 642
DB 750 GGTGCTTAATTTTCCCAATTAACCTTGCTTAACTGCTTAACTGAGAAACAGGACGAAGCT 809
QY 643 GTGATTTTCCCATTTTCCGCCAATGAAGTTAAAGTTAATAAAGAGGGAACAAAGTGA 702
DB 810 GTTCATTAACGAGCACTTACCATCGGAACCTCAAGTGAAGCGCGA-----CAATGTT 863
QY 703 CCTCAGTGGCCACTGCTGTATCATGCAATTTGGGTGAATCTTCTTATTAACAAAG 762
DB 864 CATTAAACGCTGCGTAAGCCTGCAATCAATGCTGACGGAATATTTCTCTGCAACAG 923
QY 763 CACAAGCCATGCGAGGTTGCTGCAACGCTTAAAGGGGCGAGAAATGGGTATCCT 822
DB 924 CACAGGGAATGCGGAGCGGGTGGGAGGATCACTGATTCACACGAGTGAACACCT 983
QY 823 TATGTCATCATTAACGCGCAATTTAATTTAATGGGAAACACCGTATATCCCGTC 882
DB 984 TGCTAAGTTTGCATAACGCGCAATTTATTTACTACAAACGACGCGAGAGTTCCCGCA 1043
QY 883 ATAAAGGACGCGCTATTATACAGCAGATAGATACGGCTTTAACCTTCACTGATGTC 942
DB 1044 GTCGCGCCACCCGTTATTTGATTTGATTCATGCGAGCGTTGACGCCCATC---CACGC 1100
QY 943 AGGGCAAAAGCTACCCATTTTCAGCCCAAAACCGGGTCTTGTCTCGGTGGGATGATA 1002
DB 1101 AAAAAACAGCGGTATGGTGTGACATTTACCCACTTCAGTCTGCTTTATTTCGCGACACGATA 1150
QY 1003 CCAATATTGCCAATATTCGGGTATGCTGGGAGCGGCTGCGAGCTACCCGAGCAACTG 1062
DB 1151 CTAATCTGGCAATCTCGCGCGCACCTGGAGCTCACTGACGCTTCAGGTGAGCGG 1220
QY 1063 ATAACTCCACAGGTGGGGAATTTGTTTGAATATGCAAAATCCGGAATACCAAC 1122
DB 1221 ATAAACACGCGGAGGTGGTGAACCTGTTTGAACGCTGCGTGGCTGAAGCAGATAACA 1280
QY 1123 AGCGCTACTGTCGGTGAAGATGTTCTACCAACGATGATGATCAGTTGGTGAAT 1175
DB 1281 GCCAGTGGATTCAGGTTTCGCTGCTTCCAGACTTTTACAGCAGATGCTGAT 1333

RESULT 5
US-08-910-798-1
; Sequence 1, Application US/08910798

; Patent No. 5876997
; GENERAL INFORMATION:
; APPLICANT: KRETZ
; TITLE OF INVENTION: NOVEL PHYTASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: California
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910,798
; FILING DATE: August 13, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: HAILE PH.D., LISA A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09010/029001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1272 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; IMMEDIATE SOURCE:
; CLONE: PHYTASE
; FEATURE:
; NAME/KEY:
; LOCATION:
; US-08-910-798-1

Query Match 10.8%; Score 143.6; DB 2; Length 1272;
Best Local Similarity 50.2%; Pred. No. 4.5e-38;
Matches 539; Conservative 0; Mismatches 470; Indels 64; Gaps 5;
QY 107 TAGAAGCTGTGGTATTATTTAGTCGCCATGTCGTCGCGGACCA-ACAAAGCAGC 165
DB 89 TGAAGAGTGTGGTGTGATGTCAGTCGTCATGTCGTCGTCCTCAACCAAGGCCAGCAAC 148
QY 166 TTATGAATGATGTTACGCCAGATAGTGGCCACATGCGCGTAAAGCGGGGTATTAA 225
DB 149 TGATGAGATGTCACCCAGAGCATGCGCAACCTGCGCGTAAACCTGGTGGCTGA 208
QY 226 CGCCACGTGGTGGGAGTTGGTCAATGATGGGGGGTATTTATGGTGATTAATTTCGCA 285
DB 209 CACCGCGGTTGGTGAAGTAACTCGCTATCTCGACATTTACCAAGCCAGCGCTCGGTAG 268
QY 286 GCCTTGTGTTTGGTGGG---GCGGATGTCGCCAGAGGGGGGTATATGACAGGCGAG 342
DB 269 CCGACGATGTCGGGAAAAAGGGCTGCCAGTCTGTCAGTCTCGGATTAATTGCTG 328
QY 343 ATATGATCAAGTCATCCCGCTTAAACGAGCAGGCAATTTCTGATGGTGGTCCGGGT 402
DB 329 ATGTCAGCAGCGTACCCGTAACCAAGGCGAGCA----- 363
QY 403 GTGGTTGACCGTGCATATACGCGGATTTGAAAGACCGATCCCTGTTCCATCCGG 462
DB 364 -----GATAGTCCAGTCCCGATCCGTTATTTAATCTCTC 397
QY 463 TAGAGACTGGCGTGTGTAAGTTAGACACGCAACAGATAAAGCGATTGAAGACGAT 522
DB 398 TAAAACTGCGCTTTGCCAACTGGATTAACGCGAACTGACTGACGCGATCCTCAGCAGGG 457

QY 523 TGGGCGGCGCTTAGATACGGTACGAGCGGCTACGCTAAACCTTTTGGCCAGATGGGG 582
Db 458 CAGGAGGGTCAATGTGCTGACTTACCGGCATCGGCMAACGGCGTTTGGCGAATCGAAC 517
QY 583 AGCTGCTGAATTTTGGGCGCTTCTCCCTATTGCAATCTTTGCAACAGCAAGGAAAAAGCT 642
Db 518 GGGTGTCTTATTTTCGCAATCAAACTTGTGCTTAAACGCTGAGAAACAGGACGAAGCT 577
QY 643 GTGATTTTCCGCACTTTGGGCGCAATGAAGTTAACTTAAAGAGGGAACAAAGTGA 702
Db 578 GTTCATTAACGCGAGGCAATACCATCGAACTCAAGGTGAGCGCCAGC-----AATGTCT 631
QY 703 CCTCAGTGGCGCACTGGCGTTATCATCGACATTTGGGTGAATCTTCTTATTACAAAG 762
Db 632 CATTAACCGGTGGGTAAACCTTCGATCAATGCTGACGAGATATTTCTCTGCAACAG 691
QY 763 CAACAGCCATGCGAGAGTTGCTGTCACAGCTAAAGGGGCGAGAAATTTGGGTATCCT 822
Db 692 CACAGGGAATGCGGAGCGGGTGGGAAGGATCACCGATTCAACAGTGGAAACCT 751
QY 823 TATTGTTCATACATACGCGCAATTTAATTTAATGCAAAAACACCGGTATATCGCCGTC 882
Db 752 TGTAAATTTGCAACGCGCAATTTTATTTGCTACACGCGCGGAGGTTGCCCCGA 811
QY 883 ATAAAGGAGCGCATTTATACAGCATAGATACGCTTTTAAACCTTCAACTGGATGCTC 942
Db 812 CGCGGCCACCCCGTTATTGGATTGATCATGCGAGGTTGAGCGCCATCCAC---CGC 868
QY 943 AGGGGCAAAAGCTACCCATTTCAGCCCAAAACGGGTCTGTCTCGTGGGATGATA 1002
Db 869 AAAACAGCGGTATGCTGACATTAACCCACTTCACTGTTATTGCGGACACGATA 928
QY 1003 CCAATATTGCCAATATTGCGGGTATGCTGGAGCGCACTGGCAGCTACCCGAGCAACCTG 1062
Db 929 CTATCTGCAATCTCGCGCGCACTGGAGCTCACTGGAGCTTCCCGTACGCGG 988
QY 1063 ATAACTCTACAGGTGGGATTTGTTTGAATATGGAATATGGAATATCGGAATACCAACC 1122
Db 989 ATAAACGCGCGAGGTGTGAACTGTGTGTTGAACTGTGGGTGGGTAAAGCGATAACA 1048
QY 1123 AGCGCTACCTTGGGTGAAGATGTTCTACCAACGATGATCAGTTGGTAT 1175
Db 1049 GCCAGTGGATTGAGTTTCGTGCTTCCAGACTTTACAGCAGATGGGTAT 1101

RESULT 6
US-09-489-039A-341
; Sequence 341, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489, 039A
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117, 747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 341
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-341

Query Match 4.1%; Score 53.8; DB 4; Length 1266;
Best Local Similarity 49.0%; Pred. No. 1.3e-07;
Matches 230; Conservative 0; Mismatches 232; Indels 7; Gaps 3;
QY 57 ATTGGCTGCTATTACTGCGCGTAGCGCGGCGCATCGGCTATCTTTAGACGCTGT 116
Db 66 ATCCGCTGGCGTTCAGCGCGAGACNAGGCCCGCCCGGAGGGGTATCAGCTGCAGAGT 125

QY 117 GGTATTATTGAGTCGCATCGTGTTCGTCGCCGACCAAAACAGCGCAGC-----TTATGAA 172
Db 126 GCTGATCATGAGCGCTCAACAACCTGCGTGGCCGCTTGGCCCAATAACGCGACGCTGCTGGA 185
QY 173 TGATGTTACGCCAGATTAAGTGGCCACATGCGCGTAATAAGCGGGGTATTTAACGCCACG 232
Db 186 ACATCCACCGCAAGCGCTTGGCCAGTGGAGTGTCCCGGGCGGCCAGCTGACCAACAA 245
QY 233 TGGTGGAGTTGCTCACTTATGATGGGGGGGT- TTTATGGTGTATTTCTTCGAGCGCTTG 291
Db 246 AGGCGGTGTCTGAGGTGTATATGGCCCACTATATGCGCAATGGCTGGCGCAGCAGAA 305
QY 292 GTTTGTTAGCGGG- -CGATGTCGGCAGAGGGGGGTATATGCAAGCAGATATCGA 349
Db 306 GCTGGTACACGAGCGCGAGTGGCCCGGAAACGCGGTTTATGCCCTACGCTAACACCT 365
QY 350 TCAACGTAACCGCTTAAACCGGACAGGCATTTCTTGTGTTGTGCTCCGGGGTGTGGTTT 409
Db 366 GCAGCGCACCGTCCGACACCGCGAGTTCTTTATCACCGCGCGTTCGCGGCTGGGCAT 425
QY 410 GACCGTCATATCAGCCCGATTTGAAAAGACGATCCCTCTGTTCCATCCGGTAGAGAC 469
Db 426 CCGGTGCATACACAGCTCAGATGGGCACCATGGACCCGACCTTCAACCGGTTATTAC 485
QY 470 TGGCGTGTGAAGTTAGACAAACGCAACAGATAAAGCGATTGAAGAA 518
Db 486 CGACGACTCCCGCGCTTTCGCGAGAAAGCGCTGCAGCGATGGAATAA 534

RESULT 7
US-09-489-039A-6330
; Sequence 6330, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489, 039A
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117, 747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 6330
; LENGTH: 1569
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-6330

Query Match 3.6%; Score 48; DB 4; Length 1569;
Best Local Similarity 48.0%; Pred. No. 1.4e-05;
Matches 170; Conservative 0; Mismatches 180; Indels 4; Gaps 1;

QY 73 GCGCGGTAGCGCGCAGCCATCGGCTATACCTTTAGACGCTGTGGTTATTTGAGTCGC 132
Db 364 GCGTGCATCTGCGCGCGCGGAGCTGGACCTGGAGAAAGTGTGAGCTCAGCGC 423
QY 133 CATGGTGTTCGTCGCCGA-----CCAAACACGCGCTTATGAATGATGTAGCCAGAT 188
Db 424 CACGTTATTCGTCGCCGAGCGCGCAACCGGAGACCATCGAGCGCGCCACCGCGCGA 483
QY 189 AAGTGGCCACAATGCGCGTAAAGCGGGTATTTAACGCCACGCTGGTGGAGTTGGTC 248
Db 484 CCGTGACCGAGTGACCAACCCATGACGGGAGCTCACGGCCATGCTATGCGCGCTG 543
QY 249 ACATTGATGGGGGGTTTTATGGTGTATCTTTTCGAGCGCTTGGTTTGTAGCGCGGGA 308
Db 544 GTCAACAAGGGGTGGGAAAGCCAGCAATTACCGCGAGCTCGGCTGTGTCAGGCGCGA 603
QY 309 TGTCCGCGAGAGGGGGGTATATGACAGGCGAGATATCGATCAACGCTACCGCTTAACC 368
Db 604 TGCCCGACCGCGAGTTCGATATACGTGCGCGCCAGCCCGCTGCAGCGGACCGAGGACC 663

QY 369 GGACAGGCAATTTCTTGATGGTGGCTCCGGGGTGTGGTTTACCGTGCATAT 422
Db 664 GCCCAGCGCTGGTGGATGGCGCTTCCCGGCTGCGGCTGCATTCATTAT 717

RESULT 8
US-08-920-812-23
; Sequence 23, Application US/08920812
; Patent No. 5763188
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotsugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,812
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5975 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Klebsiella pneumoniae
; STRAIN: Clinical Isolate KI-50
; US-08-920-812-23

Query Match 3.6%; Score 48; DB 1; Length 5975;
Best Local Similarity 48.0%; Pred. No. 3.3e-05;
Matches 170; Conservative 0; Mismatches 180; Indels 4; Gaps 1;
QY 73 GCGCGGTAGCGCGGAGCCATCGGGCTATCTTTAGAACGTGTGGTTATTTTTCAGTCGC 132
Db 623 GCGCTGCAATCTGCCCGCGCGGACTGGCAGCTGGGAAAGTGTGCGAGCTCAGCGGC 682
QY 133 CATGTGTTCGTCGCCGA-----CCAAACAACGAGCTTATGAATGATGTACGCCAGAT 188
Db 683 CACGGTATTCGTTCGCCCGCGGCAACCGGGAAGCCATCGAGGCGGCCACCGGCCGA 742
QY 189 AAGTCGCCACATGCGCGTAAAGCGGGTATTTAACGCCACGTGTGGAGTTTGTGTC 248
Db 743 CCGTGCAGGAGTGAACCAACCATGACGGGAGCTCACCGGCCATGCGTATCGCCCGTG 802
QY 249 ACATTGATGGGGGGTTTTATGGTGAATTAATTCGACGCTTGGTTTGTAGCGCGGGA 308

Db 803 GTCAACAAAGGCGCTGGGAGGCCAGCATTTACCGCCAGCTCGCCCTGCTGCAGGCCGA 862
QY 309 TGTCCGGCAGAGGGGGGTATATGCACAGSCAGATATCATCAACGTACCCGCTTAACC 368
Db 863 TGCCCGAGCGCGAGTCAATATAGTGGCGCCAGCCGCTGCAGCGGACCGAGCGACC 922
QY 369 GGACAGGCAATTTCTTGATGGTGGCTCCGGGGTGTGGTTTACCGTGCATAT 422
Db 923 GCCCAGCGCTGGTGGATGGCGCTTCCCGGCTGCGGCTGCATTCATTAT 976

RESULT 9
US-08-920-827-23
; Sequence 23, Application US/08920827
; Patent No. 5770375
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotsugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,827
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5975 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Klebsiella pneumoniae
; STRAIN: Clinical Isolate KI-50
; US-08-920-827-23

Query Match 3.6%; Score 48; DB 1; Length 5975;
Best Local Similarity 48.0%; Pred. No. 3.3e-05;
Matches 170; Conservative 0; Mismatches 180; Indels 4; Gaps 1;
QY 73 GCGCGGTAGCGCGGAGCCATCGGGCTATCTTTAGAACGTGTGGTTATTTTTCAGTCGC 132
Db 623 GCGCTGCAATCTGCCCGCGCGGACTGGCAGCTGGGAAAGTGTGCGAGCTCAGCGGC 682
QY 133 CATGTGTTCGTCGCCGA-----CCAAACAACGAGCTTATGAATGATGTACGCCAGAT 188
Db 683 CACGGTATTCGTTCGCCCGCGGCAACCGGGAAGCCATCGAGGCGGCCACCGGCCGA 742

QY 189 AAGTGGCCCAATGGCGGTAAGAGGGGGTATTTAAAGCCACGCTGGTGGGAGTTGGTC 248
 Db 743 CGGTGACCGAGTGGACCAACCATGACGGGAGCTCACCGGCATGGCTATGCCCGCGTG 802
 QY 249 ACATTGATGGGGGGTTTATGTTGATTAATTTCCGAGCGCTTGGTTTGTAGCGGGGA 308
 Db 803 GTCAACAAGGGGGTCCGGAAGCCAGCATACCGCCAGCTGGCTGCTGAGGCGCGGA 862
 QY 309 TGTCCCGGACAGAGGGGGGTATATGACAGGAGATATCGATCAAGTACCGCGCTTAACC 368
 Db 863 TGCCCGACGGGAGTGCATATACGTGCGCGCACCCCGCTGACGGGAGCGGAGGACC 922
 QY 369 GACAGCGCATTTCTTGATGTTGGCTCCGGGGTGGTTTACCGTGCATTAAT 422
 Db 923 GCCACGGCGCTGGTGGATGGCGCTTCCCGCGCTGCGGCTCGCTATCCATTAT 976

RESULT 10

US-08-921-177-23
 ; Sequence 23, Application US/08921177
 ; Patent No. 5798211
 ; GENERAL INFORMATION:
 ; APPLICANT: Ohno, Tsuneya
 ; APPLICANT: Matsuhisa, Akio
 ; APPLICANT: Uehara, Hirotsugu
 ; APPLICANT: Eda, Soji
 ; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
 ; NUMBER OF SEQUENCES: 25
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
 ; STREET: 6300 Sears Tower, 233 South Wacker Drive
 ; CITY: Chicago
 ; STATE: Illinois
 ; COUNTRY: United States of America
 ; ZIP: 60606-6402
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/921,177
 ; FILING DATE: 29-AUG-1997
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/362,577
 ; FILING DATE: 27-MAR-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Rin-Laures, Li-Hsien
 ; REGISTRATION NUMBER: 33,547
 ; REFERENCE/DOCKET NUMBER: 19036/32420
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 312/474-6300
 ; TELEFAX: 312/474-0448
 ; TELEX: 25-3856
 ; INFORMATION FOR SEQ ID NO: 23:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 5975 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: Genomic DNA
 ; ORIGINAL SOURCE:
 ; ORGANISM: Klebsiella pneumoniae
 ; STRAIN: Clinical Isolate KI-50
 ; US-08-921-177-23

Query Match 3.6%; Score 48; DB 1; Length 5975;
 Best Local Similarity 48.0%; Pred. No. 3.3e-05;
 Matches 170; Conservative 0; Mismatches 180; Indels 4; Gaps 1;
 QY 73 GCGCGGTAGCGCGGAGCCATCGGCTATATTTAGACAGTGGTATTTTGGAGTCG 132

Db 623 GCCTGCAATTCCTCCGCGCGGAGCTGGAGAAAGTGTGCGAGCTCAGCCGC 682
 QY 133 CATGCTGTGCTCGCCGA-----CCAAACAACGAGCTTATGATGATGTTAGCCAGAT 188
 Db 683 CACGTTATTCGTCGCGGAGCGGCAACCGGAGCCATCGAGGCGCGCACCGGCGGA 742
 QY 189 AAGTGGCCCAATGGCGGTAAGAGGGGGTATTTAAAGCCACGCTGGTGGGAGTTGGTC 248
 Db 743 CCCTGAGCCGAGTGGACCCCATGACGGGAGCTCACCGGCATGGCTATGCCCGCGTG 802
 QY 249 ACATTGATGGGGGGTTTATGTTGATTAATTTCCGAGCGCTTGGTTTGTAGCGGGGA 308
 Db 803 GTCAACAAGGGGGTCCGGAAGCCAGCATACCGCCAGCTGGCTGCTGAGGCGCGGA 862
 QY 309 TGTCCCGGACAGAGGGGGGTATATGACAGGAGATATCGATCAAGTACCGCGCTTAACC 368
 Db 863 TGCCCGACGGGAGTGCATATACGTGCGCGCACCCCGCTGACGGGAGCGGAGGACC 922
 QY 369 GACAGCGCATTTCTTGATGTTGGCTCCGGGGTGGTTTACCGTGCATTAAT 422
 Db 923 GCCACGGCGCTGGTGGATGGCGCTTCCCGCGCTGCGGCTCGCTATCCATTAT 976

RESULT 11

US-08-362-577C-23
 ; Sequence 23, Application US/08362577C
 ; Patent No. 5807673
 ; GENERAL INFORMATION:
 ; APPLICANT: Ohno, Tsuneya
 ; APPLICANT: Matsuhisa, Akio
 ; APPLICANT: Uehara, Hirotsugu
 ; APPLICANT: Eda, Soji
 ; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
 ; NUMBER OF SEQUENCES: 25
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
 ; STREET: 6300 Sears Tower, 233 South Wacker Drive
 ; CITY: Chicago
 ; STATE: Illinois
 ; COUNTRY: United States of America
 ; ZIP: 60606-6402
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/362,577C
 ; FILING DATE: 27-MAR-1995
 ; CLASSIFICATION: 536
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Rin-Laures, Li-Hsien
 ; REGISTRATION NUMBER: 33,547
 ; REFERENCE/DOCKET NUMBER: 19036/32420
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 312/474-6300
 ; TELEFAX: 312/474-0448
 ; TELEX: 25-3856
 ; INFORMATION FOR SEQ ID NO: 23:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 5975 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: Genomic DNA
 ; ORIGINAL SOURCE:
 ; ORGANISM: Klebsiella pneumoniae
 ; STRAIN: Clinical Isolate KI-50
 ; US-08-362-577C-23

Query Match 3.6%; Score 48; DB 1; Length 5975;
 Best Local Similarity 48.0%; Pred. No. 3.3e-05;

Matches 170; Conservative 0; Mismatches 180; Indels 4; Gaps 1;
QY 73 GCGCGGTAGCGCGGAGCCATCGGCTATCTTTAGAACGTGTGTTATTTTGAATGCGC 132
Db 623 GCGGTGCAATCTGCGCGCGCGGAGCTGGCAGCTGGAGAAAGTGTGAGCTCAGCGCGC 682
QY 133 CATGTTGTTCTCGCGCA-----CCAAACAACGAGCTTATGATGATGATGATGATGAT 188
Db 693 CACGTTATCTGCGCGCGCGGCGGCAACCGGGAAGCAATCGAGGCGCGCACCGCGCGGA 742
QY 189 AAGTGGCCCAATGCGCGGTAAAGCGCGGTATTTAAACGCGCACTGTTGTTAGCGGCGGA 248
Db 743 CCGTGGACCGAGTGGACCAACCATGACGCGGAGCTCACCGGCCATGCGTATGCGCGCGTG 802
QY 249 ACATTGATGCGGCGGTATTTATGCTATCTTTGCGAGCGCTTGTGTTAGCGGCGGA 308
Db 803 GTCACAAAGGCGGTGCGGAAGCGCAATTTACCGCCAGCTGCGCCTGCTGCGAGCGCGGA 862
QY 309 TGTCCGCGAGAGGGGGGTATATGACAGCGAGATATCGATCAACGTACCGCGTTAAC 368
Db 863 TGCCGACGCGGAGTGATATACGTGCGCGCGCGCGCTTCCCGGCTGCGTATCCATTAT 976

RESULT 12

US-08-920-828-23
; Sequence 23, Application US/08920828
; Patent No. 585398
; GENERAL INFORMATION:
; APPLICANT: Ohno, Teuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotsugu
; APPLICANT: Ega, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,828
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5975 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:

; ORGANISM: Klebsiella pneumoniae
; STRAIN: Clinical Isolate KI-50
US-08-920-828-23
Query Match 3.6%; Score 48; DB 2; Length 5975;
Best Local Similarity 48.0%; Pred. No. 3.3e-05;
Matches 170; Conservative 0; Mismatches 180; Indels 4; Gaps 1;
QY 73 GCGCGGTAGCGCGGAGCCATCGGCTATCTTTAGAACGTGTGTTATTTTGAATGCGC 132
Db 623 GCGGTGCAATCTGCGCGCGCGGAGCTGGCAGCTGGAGAAAGTGTGAGCTCAGCGCGC 682
QY 133 CATGTTGTTCTCGCGCA-----CCAAACAACGAGCTTATGATGATGATGATGATGAT 188
Db 693 CACGTTATCTGCGCGCGCGGCGGCAACCGGGAAGCAATCGAGGCGCGCACCGCGCGGA 742
QY 189 AAGTGGCCCAATGCGCGGTAAAGCGCGGTATTTAAACGCGCACTGTTGTTAGCGGCGGA 248
Db 743 CCGTGGACCGAGTGGACCAACCATGACGCGGAGCTCACCGGCCATGCGTATGCGCGCGTG 802
QY 249 ACATTGATGCGGCGGTATTTATGCTATCTTTGCGAGCGCTTGTGTTAGCGGCGGA 308
Db 803 GTCACAAAGGCGGTGCGGAAGCGCAATTTACCGCCAGCTGCGCCTGCTGCGAGCGCGGA 862
QY 309 TGTCCGCGAGAGGGGGGTATATGACAGCGAGATATCGATCAACGTACCGCGTTAAC 368
Db 863 TGCCGACGCGGAGTGATATACGTGCGCGCGCGCGCTTCCCGGCTGCGTATCCATTAT 976
QY 369 GGACAGGCATTTCTTGATGCTGCTCGGCGGTGTTGACCGTGCAATAA 422
Db 923 GCCCAGGCGCTGTTGATGCGCGCTTCCCGGCTGCGGCTGCTATCCATTAT 976

RESULT 13

US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELEPHONE: (703)836-9300
; TELEFAX: (703)883-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PT9pt-F18
; US-08-232-463-14

Query Match      2.6%; Score 37; DB 1; Length 7218;
Best Local Similarity 2.6%; Pred. NO. 0.2;
Matches 10; Conservative 208; Mismatches 163; Indels 0; Gaps 0;

QY 176 TGTACGCCAGTAAGTGCACATGCGCGTAAAGCGGGGTATTTAAACGCCACGTGG 235
DB 1440 TGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1391
QY 236 TCCGAGTGTGTCACATTCATCGGGGGTTCATCGTATTAATTCACGCCCTTGTTT 295
DB 1380 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1321
QY 296 GTTACGGCGGATTCGCGCAGAGGGGGGTATATGACAGGCGAGATATCGATCAACG 355
DB 1320 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1261
QY 356 TACCGCTTAACGGACAGGCAATTCCTGATGCTGCTCGGGGTGTTTGACCGT 415
DB 1260 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1201
QY 416 GCATATCAGCGCGATTTGAAAAGACCGATCCCTGTTCATCGGTAGACTGGCGT 475
DB 1200 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1141
QY 476 GTGTAAGTAGACAGCACAACAGATAAGCGATTGAAGACGATTGGCGGGCGGTT 535
DB 1140 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1081
QY 536 AGATACGGTGAGCCGCGCTA 556
DB 1080 RRRRRRRRRRRATCGCAA 1060

RESULT 14
US-09-252-991A-4062/c
; Sequence 4062, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4062
; LENGTH: 1602
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-4062

Query Match      2.6%; Score 35; DB 4; Length 1602;
Best Local Similarity 55.3%; Pred. NO. 0.36;
Matches 68; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 295 TGTACGGCGGATGTCGCGCAGAGGGGGGTATATGACAGGCGAGATATCGATCAAC 354
DB 539 TGTTCGGCGGGAACCGCGAGCGGATCGGTCGTACGGGACAGGTCGAGGTCGATCAGC 480
QY 355 GTACCGCTTAACCGACAGGCATTTCTTGATGGTGTGCTCCGGGGTGTGTTGACCG 414

; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PT9pt-F18
; US-08-232-463-14

Query Match      2.6%; Score 37; DB 1; Length 7218;
Best Local Similarity 2.6%; Pred. NO. 0.2;
Matches 10; Conservative 208; Mismatches 163; Indels 0; Gaps 0;

QY 176 TGTACGCCAGTAAGTGCACATGCGCGTAAAGCGGGGTATTTAAACGCCACGTGG 235
DB 1440 TGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1391
QY 236 TCCGAGTGTGTCACATTCATCGGGGGTTCATCGTATTAATTCACGCCCTTGTTT 295
DB 1380 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1321
QY 296 GTTACGGCGGATTCGCGCAGAGGGGGGTATATGACAGGCGAGATATCGATCAACG 355
DB 1320 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1261
QY 356 TACCGCTTAACGGACAGGCAATTCCTGATGCTGCTCGGGGTGTTTGACCGT 415
DB 1260 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1201
QY 416 GCATATCAGCGCGATTTGAAAAGACCGATCCCTGTTCATCGGTAGACTGGCGT 475
DB 1200 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1141
QY 476 GTGTAAGTAGACAGCACAACAGATAAGCGATTGAAGACGATTGGCGGGCGGTT 535
DB 1140 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1081
QY 536 AGATACGGTGAGCCGCGCTA 556
DB 1080 RRRRRRRRRRRATCGCAA 1060

RESULT 15
US-09-252-991A-4093
; Sequence 4093, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4093
; LENGTH: 1773
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-4093

Query Match      2.6%; Score 35; DB 4; Length 1773;
Best Local Similarity 55.3%; Pred. NO. 0.39;
Matches 68; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 295 TGTACGGCGGATGTCGCGCAGAGGGGGGTATATGACAGGCGAGATATCGATCAAC 354
DB 1094 TGTTCGGCGGGAACCGCGAGCGGATCGGTCGTACGGGACAGGTCGAGGTCGATCAGC 1153
QY 355 GTACCGCTTAACCGACAGGCATTTCTTGATGGTGTGCTCCGGGGTGTGTTGACCG 414
DB 1154 GTCCCGAGCCCTCGGATGCGCGCTCTCTCTCCGCGGACGCGGAGGTCGACCC 1213
QY 415 TGC 417
DB 1214 AGC 1216

Search completed: May 3, 2004, 03:58:21
Job time : 76.6484 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 3, 2004, 01:47:38 ; Search time 568.19 Seconds
(without alignments)
10548.559 Million cell updates/sec

Title: US-10-021-723B-3
Perfect score: 1325
Sequence: 1 atgtcagattagaaatcg.....aacctgcctgcacatctaa 1325

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2936184 seqs, 2261732022 residues

Total number of hits satisfying chosen parameters: 5872368

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/ECT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09C_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10C_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1325	100.0	1325	15	US-10-021-723A-3
2	1306	98.6	1326	13	US-10-282-122A-41608
3	1292.2	97.5	1325	15	US-10-021-723A-1
4	827.6	62.5	1326	15	US-10-021-723A-11
5	242	18.3	1281	15	US-10-334-672-4
6	242	18.3	1281	15	US-10-334-671-4
7	226.2	17.1	1308	15	US-10-156-660-1
8	225.2	17.0	1301	9	US-09-866-379-5
9	225.2	17.0	1301	9	US-09-866-379-6
10	225.2	17.0	1301	9	US-09-866-379-7
11	225.2	17.0	1301	15	US-10-156-660-3
12	224.6	17.0	1299	13	US-10-282-122A-7167
13	223.6	16.9	1301	9	US-09-866-379-9
14	218.8	16.5	1323	9	US-09-777-566A-1

ALIGNMENTS

RESULT 1
US-10-021-723A-3
; Sequence 3, Application US/10021723A
; Publication No. US20030101476A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Mathur, Eric
; APPLICANT: Richardson, Toby
; APPLICANT: Robertson, Dan
; APPLICANT: Barton, Nelson
; TITLE OF INVENTION: Recombinant Phytases and Uses Thereof
; FILE REFERENCE: 112766.140 (DIV-016CIP)
; CURRENT APPLICATION NUMBER: US/10/021,723A
; CURRENT FILING DATE: 2002-10-23
; PRIOR APPLICATION NUMBER: US 60/255,090
; PRIOR FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1325
; TYPE: DNA
; ORGANISM: Yersinia pestis
US-10-021-723A-3

Query Match 100.0%; Score 1325; DB 15; Length 1325;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGTCAGTATTAGAAATCGTACGGCTATCTCGGCTGTTATGATGCTTAAGCGGATTG 60
Db 1 ATGTCAGTATTAGAAATCGTACGGCTATCTCGGCTGTTATGATGCTTAAGCGGATTG 60
Qy 61 GCTGCTATTACTCGCGCGGTAGCCGCCATCTCGGCTATCTTATAGAACGTGTGGTT 120
Db 61 GCTGCTATTACTCGCGCGGTAGCCGCCATCTCGGCTATCTTATAGAACGTGTGGTT 120
Qy 121 ATTTTGATGCCCATGGTGTTCGTCGCCGACCAACACGACGCTTATGATGATGTTA 180

Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 4, Appli
Sequence 9, Appli
Sequence 1, Appli
Sequence 9, Appli
Sequence 3181, Ap
Sequence 40, Appli
Sequence 7, Appli
Sequence 5, Appli
Sequence 50, Appli
Sequence 23, Appli
Sequence 36, Appli
Sequence 301, App
Sequence 302, App
Sequence 303, App
Sequence 304, App
Sequence 25, Appli
Sequence 62, Appli
Sequence 54, Appli
Sequence 56, Appli
Sequence 29, Appli
Sequence 394, App
Sequence 394, App
Sequence 394, App
Sequence 394, App
Sequence 394, App
Sequence 394, App

```
Db 121 ATTTGAGTCGCCATGTTGTCGTCGCGACCAACAACGACGCTTATGAATGATGTTA 180
Qy 181 GCGCAGATAGTGGCCCAATAGGCGCGTAAAGCGGGTATTTAAAGCCACGCTGGTGGCG 240
Db 181 GCGCAGATAGTGGCCCAATAGGCGCGTAAAGCGGGTATTTAAAGCCACGCTGGTGGCG 240
Qy 241 AGTTGGTCACTTATGATGGGGGGGTTTATAGTGATTAATCTTTCGACGCTTGGTTGTTAG 300
Db 241 AGTTGGTCACTTATGATGGGGGGGTTTATAGTGATTAATCTTTCGACGCTTGGTTGTTAG 300
Qy 301 CCGCGGATGTCGCGCAGAGCGGGGGTATATGACAGGAGATATCGATCAACGATACCC 360
Db 301 CCGCGGATGTCGCGCAGAGCGGGGGTATATGACAGGAGATATCGATCAACGATACCC 360
Qy 361 GCTTAACCGGACAGGCAATTTCTGATGGTGGTCCCGGGTGGTGGTGGTGGTGGTGGTGGT 420
Db 361 GCTTAACCGGACAGGCAATTTCTGATGGTGGTCCCGGGTGGTGGTGGTGGTGGTGGTGGT 420
Qy 421 ATCAGGCGGATTTCAAAAGACCGATCCCTGTTCCATCCGATAGACGCTGGCGTGGTA 480
Db 421 ATCAGGCGGATTTGAAAAGACCGATCCCTGTTCCATCCGATAGACGCTGGCGTGGTA 480
Qy 481 AGTTAGACAGCACAACAGATAAAGCGATTGAAGAACGATTGGGGGGCGGTTAGATA 540
Db 481 AGTTAGACAGCACAACAGATAAAGCGATTGAAGAACGATTGGGGGGCGGTTAGATA 540
Qy 541 CCGTGAGCGGCTAGCTAAACCTTTTCCAGATGGGGCGTCTGAATTTGCGG 600
Db 541 CCGTGAGCGGCTAGCTAAACCTTTTCCAGATGGGGCGTCTGAATTTGCGG 600
Qy 601 CTTCTCCTTATGCAATCTTTGCAACAGCAAGGAAAGCGTGTGATTTGCCCACTTTG 660
Db 601 CTTCTCCTTATGCAATCTTTGCAACAGCAAGGAAAGCGTGTGATTTGCCCACTTTG 660
Qy 661 CCGCCATGAATTAAGTTAAAGAGGCAAAAGTACCTCAGTGGGCGCACTGG 720
Db 661 CCGCCATGAATTAAGTTAAAGAGGCAAAAGTACCTCAGTGGGCGCACTGG 720
Qy 721 CGTTATCATGACATTTGGTCAATCTTTTATCAAAACGCAACAGCCATGCCAGG 780
Db 721 CGTTATCATGACATTTGGTCAATCTTTTATCAAAACGCAACAGCCATGCCAGG 780
Qy 781 TTGCTGCGCAACGCTAAAGGGCGGAGATTTGGTATCTTTATTTGCTATACATACG 840
Db 781 TTGCTGCGCAACGCTAAAGGGCGGAGATTTGGTATCTTTATTTGCTATACATACG 840
Qy 841 CGCAATTTAATTAATGCAAAACACCGTATATCGCCCGTCAATAAGGACGCAATAT 900
Db 841 CGCAATTTAATTAATGCAAAACACCGTATATCGCCCGTCAATAAGGACGCAATAT 900
Qy 901 TACAGCAGATAGATACGCTTTAAACCTTCACTGGATGCTCAGGGGCAAAAGCTACCA 960
Db 901 TACAGCAGATAGATACGCTTTAAACCTTCACTGGATGCTCAGGGGCAAAAGCTACCA 960
Qy 961 TTTAGCCCAAAACCGGCTCTGTTCTCGGTGGGATGATACCAATATGCGCAATATG 1020
Db 961 TTTAGCCCAAAACCGGCTCTGTTCTCGGTGGGATGATACCAATATGCGCAATATG 1020
Qy 1021 CGGATATGCTGGGAGCGGATGCGAGTACCCGAGCAACCTGATATCTCCACCGG 1080
Db 1021 CGGATATGCTGGGAGCGGATGCGAGTACCCGAGCAACCTGATATCTCCACCGG 1080
Qy 1081 GGGGATTTGTTTGAATCTGGAATTCGGATACCCAGCGCTAGTGGCGTGA 1140
Db 1081 GGGGATTTGTTTGAATCTGGAATTCGGATACCCAGCGCTAGTGGCGTGA 1140
Qy 1141 AGATGTTCTACCAACGATGATGATGCTGATATGCGGAAATTTGGATCTGAAAATA 1200
Db 1141 AGATGTTCTACCAACGATGATGATGCTGATATGCGGAAATTTGGATCTGAAAATA 1200
Qy 1201 ACCAGCGGATATTTTCCGTTGCGATGCTGTTGTTGAAATTAACGATGACGATAGC 1260
```

```
Db 1201 ACCAGCGGATATTTTCCGTTGCGATGCTGTTGTTGAAATTAACGATGACGATAGC 1260
Qy 1261 TTTTGGAGCTGATACATTTCCAAAAGAAAGTGGCTAAGGTAATTAACCTGCTGCCACA 1320
Db 1261 TTTTGGAGCTGATACATTTCCAAAAGAAAGTGGCTAAGGTAATTAACCTGCTGCCACA 1320
Qy 1321 TCTAA 1325
Db 1321 TCTAA 1325

RESULT 2
US-10-282-122A-41608
; Sequence 41608, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 41608
; LENGTH: 1326
; TYPE: DNA
; ORGANISM: Yersinia pestis
US-10-282-122A-41608

Query Match 98.6%; Score 1306; DB 13; Length 1326;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1320; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

Qy 1 ATGTGATATTAGAAAATCGTGACGGCTATCTGGGCTGGTATTGATGCTAAGCGGATTG 60
Db 1 ATGTGATATTAGAAAATCGTGACGGCTATCTGGGCTGGTATTGATGCTAAGCGGATTG 60
Qy 61 GCTGCTATTACTGCGCGGCTAGCCCGGAGCCATCGGCTACCTTTTAAAGAACTGTTGTT 120
Db 61 GCTGCTATTACTGCGCGGCTAGCCCGGAGCCATCGGCTACCTTTTAAAGAACTGTTGTT 120
```


121 ATTTGAGTCGCGATGTTGCTCGCGACCAAC-AACGAGCTTATGATGATTT 179
121 ATTTGAGTCGCGATGTTGCTCGCGACCAACAAACGAGCTTATGATGATTT 180
180 ACGCCAGATAGTGGCCACAAATGCGCGTAAAGCGGGTATTTAAAGCGCAAGTGGTGG 239
181 ACGCCAGATAGTGGCCACAAATGCGCGTAAAGCGGGTATTTAAAGCGCAAGTGGTGG 240
240 GAGTGGTCACATGATGGGGGGTTTTATGGTGATATCTTTCGACGCTTGGTTGTTA 299
241 GAGTGGTCACATGATGGGGGGTTTTATGGTGATATCTTTCGACGCTTGGTTGTTA 300
300 GCGGCGGGATGTCGCGCAGAGGGGGGTATATGCAAGCGAGATATCGATCAACGTACC 359
301 GCGGCGGGATGTCGCGCAGAGGGGGGTATATGCAAGCGAGATATCGATCAACGTACC 360
360 CGCTTAACCGGACAGGCAATTTCTTGATGTTGGCTCCGGGGTGTGGTTTGAACGTCAT 419
361 CGCTTAACCGGACAGGCAATTTCTTGATGTTGGCTCCGGGGTGTGGTTTGAACGTCAT 420
420 AATCAGCGCCGATTTGAAAGAGCGATCCCTCTGTTCCATCCGCTAGAGACTGCGGTGCT 479
421 AATCAGCGCCGATTTGAAAGAGCGATCCCTCTGTTCCATCCGCTAGAGACTGCGGTGCT 480
480 AAGTTAGACAAACGACAAACAGATTAAGCGATTAAGCGATTAAGCGATTAAGCGATTA 539
481 AAGTTAGACAAACGACAAACAGATTAAGCGATTAAGCGATTAAGCGATTAAGCGATTA 540
540 ACGGTAGCGAGCGCTACGCTTAACCTTTTGGCCAGATGGGACGCTGCTGATTTTGG 599
541 ACGGTAGCGAGCGCTACGCTTAACCTTTTGGCCAGATGGGACGCTGCTGATTTTGG 600
600 GCTTCTCTTATTTGCAATCTTTGCAACAGCAAGGAAACAGTGTGATTTTGGCCACTTT 659
601 GCTTCTCTTATTTGCAATCTTTGCAACAGCAAGGAAACAGTGTGATTTTGGCCACTTT 660
660 GCGGCCAATGAGTAAAGTAAAGAGGAGCAAAAGTGAACCTCTGAGTGGGCGCACATG 719
661 GCGGCCAATGAGTAAAGTAAAGAGGAGCAAAAGTGAACCTCTGAGTGGGCGCACATG 720
720 GCGTTATCATCGATGCTGCTGAAATCTTCTTATTACAAACGCAAGCCATGCGCAGAG 779
721 GCGTTATCATCGATGCTGCTGAAATCTTCTTATTACAAACGCAAGCCATGCGCAGAG 780
780 GTTGCCTGGCAACGGCTAAAGGGGCGGAGAAATGGGTATCCTTATTTGCTATACATAAC 839
781 GTTGCCTGGCAACGGCTAAAGGGGCGGAGAAATGGGTATCCTTATTTGCTATACATAAC 840
840 GCGCAATTTAATTTAATGCAAAACACCGCTATATCGCCGTCATTAAGGGAGCGCCATTA 899
841 GCGCAATTTAATTTAATGCAAAACACCGCTATATCGCCGTCATTAAGGGAGCGCCATTA 900
900 TTACAGAGATAGATAGCGCTTTAACCTTTCAACTGGATGCTCAGGGGCAAAAGTACCC 959
901 TTACAGAGATAGATAGCGCTTTAACCTTTCAACTGGATGCTCAGGGGCAAAAGTACCC 960
960 ATTTACGCCAAACCGGGTCTTGTCTCGGTGGCATGATACCAATATGCGCAATATT 1019
961 ATTTACGCCAAACCGGGTCTTGTCTCGGTGGCATGATACCAATATGCGCAATATT 1020
1020 GCGGGTATGCTGGAGCGCACTGGCACTACCGAGCAACCTGATATATCTCACCGAGT 1079
1021 GCGGGTATGCTGGAGCGCACTGGCACTACCGAGCAACCTGATATATCTCACCGAGT 1080
1080 GGGGGATTTGTTTGAACATGCGCAAAATCCGGATAACCAAGCGCTACGTTGGGTG 1139
1081 GGGGGATTTGTTTGAACATGCGCAAAATCCGGATAACCAAGCGCTACGTTGGGTG 1140
1140 AAGATGTTCTACCAACGATGGATCAGTTCGGTAAATGCGGAAATTTGGATCTGAAAAAT 1199
1141 AAGATGTTCTACCAACGATGGATCAGTTCGGTAAATGCGGAAATTTGGATCTGAAAAAT 1200
1200 AACCCAGCGGGTATTATTTCCGTTGAGTTGCTGCTGTTGTTGTTGTTGTTGTTGTTGTT 1259

1201 AACCCAGCGGTATTTTCGTTGCGAGTCTCTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1260
1260 CTTTTCGAGCTTGATGATCAATTCCTCAAAAGAAAGTGGCTTAAGGTAAATGAACTGCTGCCAC 1319
1261 CTTTTCGAGCTTGATGATCAATTCCTCAAAAGAAAGTGGCTTAAGGTAAATGAACTGCTGCCAC 1320
1320 ATCTAA 1325
1321 ATCTAA 1326

RESULT 3
US-10-021-723A-1
; Sequence 1, Application US/10021723A
; Publication No. US20030101476A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Mathur, Eric
; APPLICANT: Richardson, Toby
; APPLICANT: Robertson, Dan
; APPLICANT: Barton, Nelson
; TITLE OF INVENTION: Recombinant Phytases and Uses Thereof
; FILE REFERENCE: 112766.140 (DIV-016CIP)
; CURRENT APPLICATION NUMBER: US/10/021.723A
; CURRENT FILING DATE: 2002-10-23
; PRIOR APPLICATION NUMBER: US 60/255,090
; PRIOR FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1325
; TYPE: DNA
; ORGANISM: Yersinia pestis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 470..487, 491, 492, 522
; OTHER INFORMATION: n = A,T,C or G
US-10-021-723A-1

Query Match 97.5%; Score 1292.2; DB 15; Length 1325;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 1316; Conservative 0; Mismatches 8; Indels 2; Gaps 2;

QY 1 ATGTCAGTATTAGAAAATCGTGTACGGCTATCTGGGCTGGTATTGATGCTTAAGCGGATTG 60
DB 1 ATGTCAGTATTAGAAAATCGTGTACGGCTATCTGGGCTGGTATTGATGCTTAAGCGGATTG 60
QY 61 GCTGCTATTACTGCGCGGTAGCCGCGAGCCATCGGCTATATCTTTAGAACGTTGTGTT 120
DB 61 GCTGCTATTACTGCGCGGTAGCCGCGAGCCATCGGCTATATCTTTAGAACGTTGTGTT 120
QY 121 ATTTTGTAGTCGCGCATGGTGTTCGCTCGCCGACCAAC-AACGAGCTTATGAATGATGTT 179
DB 121 ATTTTGTAGTCGCGCATGGTGTTCGCTCGCCGACCAACCAACCAACGAGCTTATGAATGATGTT 180
QY 180 ACGCCAGATAAGTGGCCAAATGCGCGTAAAGCGGGTATTTAAAGCGCAAGTGGTGG 239
DB 181 ACGCCAGATAAGTGGCCAAATGCGCGTAAAGCGGGTATTTAAAGCGCAAGTGGTGG 240
QY 240 GAGTGGTCAATGATGGGGGGTTTTATGGTGATATCTTTCGACGCTTGGTTGTTA 299
DB 241 GAGTGGTCAATGATGGGGGGTTTTATGGTGATATCTTTCGACGCTTGGTTGTTA 300
QY 300 GCGGCGGGATGTCGCGCAGAGGGGGGTATATGCAAGCGAGATATCGATCAACGTACC 359
DB 301 GCGGCGGGATGTCGCGCAGAGGGGGGTATATGCAAGCGAGATATCGATCAACGTACC 360
QY 360 CGCTTAACCGGACAGGCAATTTCTTGATGTTGGCTCCGGGGTGTGGTTTGAACGTCAT 419
DB 361 CGCTTAACCGGACAGGCAATTTCTTGATGTTGGCTCCGGGGTGTGGTTTGAACGTCAT 420
QY 420 AATCAGCGCGATTTGAAAGAGCGATCCCTCTGTTCCATCCGCTAGAGACTGCGGTGCT 479

Db 421 AATCAGGCGATTGAAAAGACCGATCCCTGTTCCATCCGTTAGAGTGGCGTGT 480
QY 480 AAGTTAGCAACGCAACAAAGATGAAGATGAAGAACGATTGGCGGCGCGTTAGAT 539
Db 481 AAGTTANACGNNGCACAAACACGATGAAGCGATTGAAGAACANTTGGCGGCGCGTTAGAT 540
QY 540 ACGGTGAGCGAGCGTACGCTAAACCTTTTCCAGATGGGAGCGTGTCTGAATTTGCG 599
Db 541 ACGGTGAGCGAGCGTACGCTAAACCTTTTCCAGATGGGAGCGTGTCTGAATTTGCG 600
QY 600 GCTTCTCTTATTTGCAATCTTTGCAACGAGCAAGGAAACGTTGATTTTCCACTTT 659
Db 601 GCTTCTCTTATTTGCAATCTTTGCAACGAGCAAGGAAACGTTGATTTTGC-CACTTT 659
QY 660 GCGGCAATTAATTTAATGCAAAACACCGTATATCGCCCTGATAAGGAGCGCCATTA 899
Db 840 GCGCAATTTAATTTAATGCAAAACACCGTATATCGCCCTGATAAGGAGCGCCATTA 899
QY 900 TTACAGCGATAGATACGCTTTTAACTTCACTGGATGCTCAGGGGCAAGCTACCC 959
Db 900 TTACAGCGATAGATACGCTTTTAACTTCACTGGATGCTCAGGGGCAAGCTACCC 959
QY 960 ATTTGAGCCCAAAACCGGCTCTTTGTTCTCGTGGGCGATGATACCAATTTGCCAATTT 1019
Db 960 ATTTGAGCCCAAAACCGGCTCTTTGTTCTCGTGGGCGATGATACCAATTTGCCAATTT 1019
QY 1020 GCGGTATGCTGGGAGCGGACGTCGAGCTACCGGAGCAACCTGATATATCTCCACGAT 1079
Db 1020 GCGGTATGCTGGGAGCGGACGTCGAGCTACCGGAGCAACCTGATATATCTCCACGAT 1079
QY 1080 GCGGATGCTGTTTGAACATGCGAATATGCGAATATGCGAATATGCGAATATGCGGTTG 1139
Db 1080 GCGGATGCTGTTTGAACATGCGAATATGCGAATATGCGAATATGCGGTTG 1139
QY 1140 AAGATGTTCTACCAACGATGATCAGTTGCGTAAATGCGAATAACCGGCTAGCTGAAAAT 1199
Db 1140 AAGATGTTCTACCAACGATGATCAGTTGCGTAAATGCGAATAACCGGCTAGCTGAAAAT 1199
QY 1200 AACCAGCGGATTTATTTCCGTTGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1259
Db 1200 AACCAGCGGATTTATTTCCGTTGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1259
QY 1260 CTTTGGAGCTTGATACATTCGAAAGAAAGTGGCTAAGTAAATTTGAACCTGCTGCCAC 1319
Db 1260 CTTTGGAGCTTGATACATTCGAAAGAAAGTGGCTAAGTAAATTTGAACCTGCTGCCAC 1319
QY 1320 ATCTAA 1325
Db 1320 ATCTAA 1325

RESULT 4
US-10-021-723A-11
; Sequence 11, Application US/10021723A
; Publication No. US20030101476A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Mathur, Eric
; APPLICANT: Richardson, Toby
; APPLICANT: Robertson, Dan

; APPLICANT: Barton, Nelson
; TITLE OF INVENTION: Recombinant Phytases and Uses Thereof
; FILE REFERENCE: 112766.140 (DIV-016CIP)
; CURRENT APPLICATION NUMBER: US/10/021,723A
; PRIOR FILING DATE: 2002-10-23
; PRIOR APPLICATION NUMBER: US 60/255,090
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1326
; TYPE: DNA
; ORGANISM: Versinia pestis
US-10-021-723A-11

Query Match 62.5%; Score 827.6; DB 15; Length 1326;
Best Local Similarity 77.0%; Pred. No. 5.9e-261;
Matches 1021; Conservative 0; Mismatches 304; Indels 1; Gaps 1;

QY 1 ATGTCAGTATTAGAAATCGTGTACGCTATCTGGCTATCTGGCTATGATGCTAAGCGGATG 60
Db 1 ATGTCGGAGCTGAGAACCGCTCGCCCTTCCGGTTTAGTGTAAATGCTGTGCGGCTG 60
QY 61 GCTGCTATTACTGCGCGGCTAGCCCGGAGCATGGGCTATACCTTAGAAGCTGTGTT 120
Db 61 GCTGCTATCACCGCGCTGTGGCCCGCAACCATCGGGTACACCTTAGAAGCTGTGTT 120
QY 121 ATTTTGTAGTCGCTATGCTGCTCGCGCAACAAACA-ACGACGCTTATGAATGATGTT 179
Db 121 ATTTTGTAGTCGCTATGCTGCTCGCGCAACAAACA-ACGACGCTTATGAATGATGTT 179
QY 180 ACGCCAGATAGTGGCCACATGGCCCGTAAAGGGGGTATTAACGCCACGTTGGTGG 239
Db 181 ACACCTGATAAGTGGCTCAGTGGCGGTTAAAGCGGCTATTTGACTCTCTCGTGGCG 240
QY 240 GAGTTCGTCACATTGATGGGGGTTTATGCTGATTAATCTTTCGACGCTTGGTTGTTA 299
Db 241 GACTGCTCACCTGATGGCGGTTCTATGGCGATTATTTCCGACAGTTTGGGCTCTTTG 300
QY 300 GCGGCGGATGTCGCGCAGAGGGGGGTATATGACAGCAGATATCGATCAACGTCAC 359
Db 301 GCGCGCGGCTGCGCGCAGAGGGCGGTATATGACAGCAGATATCGACGCGCACT 360
QY 360 CGCTTAAACCGGACAGGCACTTTCTTGTATGCTGTGGCTCCGGGCTGTGTTTGAACG 419
Db 361 CGTTTAAACCGGACAGGCTTTCTTGTATGCTGTGGCGCGGTTGCGGCTGACTGCTC 420
QY 420 AATCAGCGGATTTGAAAAGACCGATCCCTGTTCCATCCGTTAGAGACTGGCGTGTGT 479
Db 421 AATCAGCGGATTTGAAAAGACCGATCCCTGTTCCATCCGTTGAAAACCGCGCTCTGT 480
QY 480 AAGTTAGCAACGCAACAAAGATGAAGATGAAGAACGATTGGCGGCGCGTTAGAT 539
Db 481 AACTGCAACACCGCCCAACCGGATAGGCAATTTAGGAAACGCTGGGCGGCGGCTTAGAC 540
QY 540 ACGGTGAGCGAGCGCTACGCTAAACCTTTTGGCCAGATGGGGAACGCTGTGAATTTGCG 599
Db 541 ACGGTAAAGCGAGCGCTATGCAAAACCGTTTGGCAAAATGGGCGATGCTCCTGAACTTCG 600
QY 600 GCTTCTCTTATTTGCAATCTTTGCAACGAGCAAGGAAACGTTGATTTTGCCTCCACTTT 659
Db 601 GCGAGTCGCTACTGCAAGTCACTGAGCAGAGGGGAAACCTTGTACTTCGACACTTTT 660
QY 660 GCGGCAATTAATTTAATGCAAAACACCGTATATCGCCCTGATAAGGAGCGCCATTA 719
Db 661 GCGGCAACGAACTTAATGTAACAGGAAAGCACGAAAGTTACCTGTGAGGCGGCGCTG 720
QY 720 GCGTTATCATCGATGCTGGTGAATCTTTTATTAACAAACGCAACGCAATGCCAGAG 779
Db 721 GCGCTGTCTAGCACGTTGGCGGAAATCTTTCTTGTGCAAGCGCGGCGATGCGCCGAA 780
QY 780 GTTGCCTGCGCAACGCTTAAAGGGGCGGAGATTTGGTATCTCTTATTTGCTATACATAAC 839

Db 781 GTAGCGTGGCAGCGTTTGAAGCGCGCTGAGAACTGGGTGCTCTTCTTGAGCGCTGCACAAT 840
 Qy 840 GCGCAATTAATTAATGCAAAACACCGTATATCGCCGCTAATAAGGAGCGCAATA 899
 Db 841 GCACAGTTCAACCTGATGCTAAACCGCATATCATGTACGCGCCACAAAGGCAACGCGTT 900
 Qy 900 TTACACAGATAGATACGCGCTTTAAACCTTCAACTGGATGCTTCAGGGGCAAAAGCTACCC 959
 Db 901 TTACAGCAATCGATACGCACTGACCTTGCACCTGGACGCGCCAGGGGCAAAATCTGCG 960
 Qy 960 ATTTACAGCCAAACACCGGCTTTGCTTCTCGGTGGGCAATGATACCAATATTCGCAATAT 1019
 Db 961 ATCTCGGCTCAGAACCGTGTATTTATCTCGGGTGGCCAGCACCAATATTTGCTAACATC 1020
 Qy 1020 GCGGGTATCTCGGGAGCGCACTGGCAGTACCGGAGCACTGATTAATCTCCACAGGT 1079
 Db 1021 GCGGGTATCTCGGGAGCGCACTGGCAGTACCGGAGCACTGATTAATCTCCACAGGT 1080
 Qy 1080 GCGGGATGCTTTTGAATATGCGCAAAATCCGGATAACCAACAGCGCTACGTTGCGGTG 1139
 Db 1081 GCGGGTCTGCTTTTGAATATGCGCAAAATCCGGATAACCAACAGCGCTACGTTGCGGTG 1140
 Qy 1140 AGAGTCTTACCAACAGTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1199
 Db 1141 AGAGTCTTATCAGACCATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
 Qy 1200 AACCCAGCGGTATTTCCGTTGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1259
 Db 1201 AATCCGCGCGCATCATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
 Qy 1260 CTTTGGAGCTGATACATTCAGAGAAAGTGGCTAAGTAAATGAACTGCTGCCAC 1319
 Db 1261 CTTTGGAGCTGATACATTCAGAGAAAGTGGCTAAGTAAATGAACTGCTGCCAC 1320
 Qy 1320 ATCTAA 1325
 Db 1321 ATTTAA 1326

RESULT 5
 US-10-334-672-4
 ; Sequence 4, Application US/10334672
 ; Publication No. US20030157646A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lanahan, Mike
 ; APPLICANT: Koepf, Edward
 ; APPLICANT: Kretz, Keith
 ; TITLE OF INVENTION: Microbially-Expressed Thermotolerant Phytase For Animal Feed
 ; FILE REFERENCE: SYNG-P01-001
 ; CURRENT APPLICATION NUMBER: US/10/334,672
 ; PRIORITY FILING DATE: 2002-12-30
 ; PRIOR APPLICATION NUMBER: 60/344,523
 ; PRIORITY FILING DATE: 2001-12-28
 ; NUMBER OF SEQ ID NOS: 4
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 4
 ; LENGTH: 1281
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: No. US20030157646A19X gene
 US-10-334-672-4

Query Match 18.3%; Score 242; DB 15; Length 1281;
 Best Local Similarity 52.5%; Pred. No. 3.7e-68;
 Matches 628; Conservative 0; Mismatches 555; Indels 13; Gaps 4;
 Qy 93 ATCGGGTATATCTTTAGAACGTTGGTATTTTGGATCGCCATGTTGCTCGCCGAC 152
 Db 39 ACCAGAAATTGAAGTTGAATCTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 98
 Qy 153 CAA-ACACGAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 211

Db 99 TAAGGCTACTCAGTTGATGCAAGATGTTACTCCAGATGCTTGGCTACTCTACCTGGCTGTAA 158
 Qy 212 AGCGGGTATTTAACGCCACGTCGTCGCGAGTTGGTCAATTCATGATGGGGGGTATTTATGG 271
 Db 159 GTTGGGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 218
 Qy 272 TGATTACTTTCCGAGCCCTTGGTTGTTT---AGCGGGGAGATGTCGCGCAGAGGGGGGGT 328
 Db 219 ACAAAGATTGGTTCGCTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 278
 Qy 329 ATATGCAAGCGAGATATCGATCAACGTCACCGCTTAACCGGACAGGCATTTCTTGATGG 388
 Db 279 TGTCTATCATTTGCTGATTTGATGAAAGAACTAGAAAGAACTGGTGAAGCTTCGCTGCGCG 338
 Qy 389 TGTGGCTCCCGGGTGTGGTTTGACCGGTGATTAATCAGGCCGATTTGAAAAAGACCGATCC 448
 Db 339 TTTGGCCCAAGACTGTGCTATCACTGTTCACATCAAGCTGATATTTCTCTCCAGATCC 398
 Qy 449 CTTGTTCCATCCGCTGAGAGACTGCGGTGTGTAGTTAGACAAACGACAGATTAAGC 508
 Db 399 ATTGTTCAACCCATTTGAAGAACTGCTGTCTGCTCAATTTGATAACGCTAACCTTACTGATGC 458
 Qy 509 GATTGAAGAACGATTTGGCGGCGCTTGTAGATACGGTGAGCCAGCGCTTACGCTAAACCTTT 568
 Db 459 CATCTTGAAGAGCTGTGCTGCTATCGCTGACTTCACTGGTCACTTACCAGATCGCTT 518
 Qy 569 TGCCAGATGGGGAACGTCGTAATTTTGGCGCTTCTCTTATTTGCAAACTCTTTGCAACA 628
 Db 519 CAGAGAAATTGGAAGAGCTTGAACCTTCCCAACAATTAATCTGTTTGAAGAGAGAGAA 578
 Qy 629 GCAGGAAACACGCTGATTTTGCCTTTCGCGCAATGAAAGTAAAGTAAATTAAGA 688
 Db 579 GCAAGACGAATCTTTGCTTTCCTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 638
 Qy 689 AGGACAAAAGTACCTCAGTGGGCACTGGGCTTATCATGACATTTGGGTGAATCTTT 748
 Db 639 TTG-----TGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 692
 Qy 749 CTTATTACAAACGCAAGCCATGCGAGAGTTGCTGCGCAACGGCTTAAGGGCGGA 808
 Db 693 CTTGTTGCAACAGCTCAAGGTATGCGAGAACCGAGTTGGGTGAGATCACTGATTTCTCA 752
 Qy 809 GAATTTGGGTATCTTTTATTGCTTATTAACGCGCAATTTAATTAATGCAAAACACC 868
 Db 753 CCAATGGAACACCTTTGTTGCTTTCGCAACGCTCAATTCGATTTGCTGCAAGAACTCC 812
 Qy 869 GTATATCGCCGCTATTAAGGAGCGCATTTATTAACAGCAGATAGATACGCTTTTAAACCT 928
 Db 813 AGAAGTCCGTAGATCCAGAGCTACTCCATTTGTTGACTTGTATCAAGACCGCTTTGAC--- 869
 Qy 929 TCAACTGGATGCTCAGGGGCAAAAGCTACCCATTTCAAGCCCAAAACCGGGTCTTTGCTCT 988
 Db 870 TCCACCCCAACAGAGCAAGCTTACGCTGTTTACCTTGCCTTCTGCTGTTTCTCAT 929
 Qy 989 CGGTGGGATGATACCAATATTTGCAATATTTGGGGTATGCTGGGAGCGAGCTGGCAGCT 1048
 Db 930 TGCCTGTCAGATATCACTTGGCTAACTTGGGTGCTGCTTGGAAATGGAATCGACCTT 989
 Qy 1049 ACCGAGCAACCTGATTAATCTCCACAGGTGGGGGATTTGGTTTGAACATATTCGCAAAA 1108
 Db 990 GCGAGTTCAACAGATTAACCTCCACAGGTGGTGAATTTGGTCTTCGAAAGATGCGCTCG 1049
 Qy 1109 TCCGATTAACCAACAGCGCTTACGTTGGGTGAAGATGTTCTTCAACAGATGATGATGATGATGAT 1168
 Db 1050 ACTGCTGATTAATCTCAATGGAATTCAGTCTCCTGCTTCTTCCAAACCTTTGCAACAAAT 1109
 Qy 1169 GCGTAATCCCGAAATTTGATCTGAAATAAATCCAGCGGGTATTTTCCGTTGCGAT 1228
 Db 1110 GAGAGCAAGACTCCTATTTGCTTGAACACTCCACAGGTGAAGTCAAGTCAAGTCAAGTCAAGT 1169
 Qy 1229 TGCTGGTTGTGAAATAACCGTGAAGTAAAGCTTTGGAGCTTTGGAGCTTTGATCAATTTCCAAA 1284
 Db 1170 TGGTTGTGAAGAGAAAGAAAGCTCAAGGTATGTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1225

RESULT 6

```

US-10-334-671-4
; Sequence 4, Application US/10334671
; Publication No. US20030170293A1
; GENERAL INFORMATION:
; APPLICANT: Laranhan, Mike
; APPLICANT: Kreutz, Edward
; APPLICANT: Kreutz, Keith
; TITLE OF INVENTION: Microbially-Ex
; FILE REFERENCE: 70098
; CURRENT APPLICATION NUMBER: US/10/
; CURRENT FILING DATE: 2002-12-30
; PRIOR APPLICATION NUMBER: 60/344, 5
; PRIOR FILING DATE: 2001-12-28
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 1281
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
;
; -OTHER INFORMATION: No. US20030170
US-10-334-671-4

```

Query Match	18.3%; Score 242; DB 15; Length 1281;
Best Local Similarity	52.5%; Pred. No. 3.7e-68;
Matches	628; Conservative 0; Mismatches 555; Indels 13; Gaps 4;
QY	93 ATCGGGCTATACTTTAGAACCTGTGGTTATTTTGTAGTCGCCATGTGTTTCGCTGCCGCAC 152
Db	39 ACCAGAAATTGAAGTTGGAACTGTGTGCATGTCTCCAGACACGGTGTGTAGAGCTCCAAC 98
QY	153 CAA-AACACGCGAGCTTATGAATGATGTTACGCCAGATAAAGTGCCACAAATGCCCGTAAA 211
Db	99 TAAGGCTACTCAGTTGATGCAAGATGTTACTCCAGATGCTTGGCTTACCTGGCCCTGTATA 158
QY	212 ACGGGGTATTTAAGCCACCTGTGTGGGAGTTGTGCATTTGATGGGGGGGGTTTATGG 271
Db	159 GTTGGGTGAATTGACTCCAGAGGTGTGTAATTGATTGTCTTCTGGGTCACTACTGGAG 218
QY	272 TGATTACTTTCCGACCTTGGTTTGTGT - - - AGCGCGGGATGTCCGGCAGAGGGGGGGGT 328
Db	219 ACAAGATTGGTTGCTGATGGTTTGTGCCAAAGTGTGGTTGTCCACAATCTGTGTCAAAGT 278
QY	329 ATATGCACAGCAGATATCGATCAACGTACCGCTTAAACGGACAGGCATTTCTTTGATGG 388
Db	279 TGCTATCATTTGCTGATGTGTGATGAAGAAATAGAAAGACTGTGTGAAGCCCTTCGCTGCCGG 338
QY	389 TGTGCTCCGGGGTGTGGTTTTCACCGTGCATAATCAAGCGCGATTTTGAAAAGACACGATCC 448
Db	339 TTTGGCCCCAGACTGTGTCTACTCTGTTTCACTCAAGCTGATACTTCTCTCCAGATCC 398
QY	449 CTTGTTTCATCCGGTAGAGACTGGCGTGTGTAGTTAGACAACGCAACAACAGATAAAGC 508
Db	399 ATGTGTTCAACCCATTGAAGACTGTGTGTCTGTCAATTGGATTAACGCTTAACGTTTACTGTATGC 458
QY	509 GATTGAAGAACGATTGGCGGGCGGTTAGATACGGTGGAGCCGCTACGCTTAAACCTTT 568
Db	459 CATCTTGGAAAGAGCTGGTGGTCTATCGCTGACTTCACTGTGTCATCAACAACCTGCCTT 518
QY	569 TGCCACAGATGGGGGAGCTGCTGAATTTTGGCGCTTCTCTTTATTGGAATCTTTGCAACA 628
Db	519 CAGAGAAATTGAAAGAGCTTTGAACCTTTCCACAACTCTAACTTGTGTGTTTGAAGAGAGAGAA 578
QY	629 GCAAGGAAAAACGTGTGATTTTGCCTACTTTGGCGGCAATGAAGTTTAAAGTTAAATAAGA 688
Db	579 GCAGACGAACTTTGTTCCCTTGACTCAAGCCTTGGCACTTGAAATTTGAAAGTCTCTGCTGA 638
QY	689 AGGGACAAAAGTAGACCCCTCACTGGGCGCACTGGCGGTATATCATGACATTTGGGTGAAATCTT 748
Db	639 TTG-----TGCTCTCTTGACTGGTGTGTCTCTTGGCTTCTATGTCACATGAATCTT 692

; FEATURE:
 ; OTHER INFORMATION: modified phytase enzyme
 ;
 ; NAME/KEY: CDS
 ; LOCATION: (1)...(1308)
 US-10-156-660-1

Query Match 17.1%; Score 226.2; DB 15; Length 1308;
 Best Local Similarity 53.3%; Pred. No. 5.9e-63;
 Matches 572; Conservative 0; Mismatches 488; Indels 13; Gaps 4;

QY 107 TAGAAGCTGTGGTATTTTGGTGCAGTGTGTTGGTCCGCGACCAA-ACAACGAGC 165
 DB TGGAAAGTGTGGTGAATGTTCAGTGTGATGGTGTGGTCTCCACCAAGCCAGCAAC 148
 QY 166 TTATGAATGATGTACCGCAGATAGTGCACATAGTCCGATGAAACCGGGGTATTAA 225
 DB TGAATGAGATGTGACCCAGACGATGCGCCAACTGGCCCGGTAAATCGGGTGAAGTGA 208
 QY 226 CGCACGCTGTGGAGTGTGTCATATGATGGGGGGTATTTATGCTGATTACTTTGCA 285
 DB CACCGCGGTGTGGTGAATATCGCTATCTCGACATTAATGCGGTCTGGTAG 268
 QY 286 GCCTTGTGTTAGC---GGCGGATGTCCGAGAGGGGGGTATATGCAAGGAG 342
 DB CGACCGATGTGCTGCTAAATATGCTGCCCGACCTCTGGTCAAGTCCGATATTGCTG 328
 QY 343 ATATGATCAAGTACCCCGCTTAAACGGAGACGATTTCTTGATGGTGTGCTCCGGGT 402
 DB ATGTGACGAGGTATCCCGTAAACAGCGAGCCCTTCGCGCGGGCTGCACCTGACT 388
 QY 403 GTGGTTGACGTGCAATATCAGCGGATTTGAAAGAACCGATCCCTGTTCCATCCGG 462
 DB GTGCAATACCGTACATACCCAGGAGATACGTCCAGTCCCGATCCGTTATTTAATCCTC 448
 QY 463 TAGAGACTGGCTGTGTAAAGTTAGACAACGCAACAGATAAAGCGATTTGAAGACGAT 522
 DB TAAAGACTGGCTTTGCCAATGAGTACGGAACGTGACTGACGCGATCTTCAGAGGG 508
 QY 523 TGGCGCGCGGTAGATACGTTGACGACGCGGTACGCTAAACCTTTTCCAGATGGGG 582
 DB CAGGAGGGTCAATTGCTGACTTTTACCGGCAATATCAAAACGCGGTTTCGCAACTGGAAC 568
 QY 583 ACGTGTGAATTTGGCGCTTCTCTTATGCAATCTTTGCAACAGCAAGGAAACGTT 642
 DB GGGTGTCTAATTTTCGCAATCAACTGTGCTTTAAACGTGAGAAACAGGACGAAAGCT 628
 QY 643 GTGATTTGCCACTTTGGGCGCAATGAAGTTAAAGTTAAATAAGAGGGACAAAGTGA 702
 DB GTTCATTAAACGAGGCAATPACCATCGGAATCAAGGTGAGCGGCTGTCTCTCATTA 688
 QY 703 CCCTCAGTGGCCACTTGGCGTTATCATGACATTTGGGTGAATCTTCTTATTAACAACG 762
 DB CC-----GGTGGGTAGCCTCGCATCATGCTGACGAGATATTTCTCTGACAAAG 742
 QY 763 CAACAGCATGCGAGGTGCTTGGCAACGCGCTAAAGGGGCGGAGAAATTTGGGTATCCT 822
 DB CACAGGGAATGCGGAGCGGGGTGGGAAGGATCACCGATTCACCAAGTGAACACCT 802
 QY 823 TATGTCTATTACATACGCGCAATTTAATTAATGCGCAAAACACCGTATATCCCGGTC 882
 DB TGTAAATTTGATACGCGCAATTTGATTTGCTCAACGACGCGGAGGTTTCCCGCA 862
 QY 883 ATAAAGGAGCGCAATTTATPACGAGATAGATACGGCTTTAAACCTTCAACTGGATGCTC 942
 DB GCGCGCCCAACCCCGTTATTAGATTTGATCAAGACAGCGTTTACGCGCCCATC---CACCGC 919
 QY 943 AGGGGCAAAAGTACTCCATTTTCAGCCCAAAACCGGCTCTTCTCTGGTGGGATGATA 1002
 DB AAAAAAGCGGTATGCTGTGATCAATTCACCTTTCAGTGTCTGTTATCGCCGGAACCATTA 979
 QY 1003 CCAATATGCCAATATTCGGGTATGCTGGGAGCGCACTGGCAGTACCCGAGCAACCTG 1062

Db 980 CTAATCTGGCAAAATCTCGCGCGCACTGGAGCTCAACTGGACGCTTCCCGGTGAGCGG 103
 QY 1063 ATAATACTCCACAGGTGGGGATTTGGTTTTTGAACATATGCAAAATCCGGATAACCAACC 1122
 Db 1040 ATAACACGCGCGCAGTGTGTAACCTGTTTGAACGCTGGCGTCCGCTAAGCATTAACA 1099
 QY 1123 ACGGTACGTTGCGGTGAAGATGTTCTACCAAGCATGATGATTCAGTTCGGTAAAT 1175
 Db 1100 GCAGTGGATTGAGTTTGGCTGTCTTCCAGACTTTTACAGCAGATGCGTAT 1152

RESULT 8
 US-09-866-379-5
 ; Sequence 5, Application US/09866379
 ; Patent No. US20020136754A1
 ; GENERAL INFORMATION:
 ; APPLICANT: DIVERSA CORPORATION
 ; APPLICANT: SHORT, Jay
 ; APPLICANT: KRETZ, Keith
 ; APPLICANT: GRAY, Kevin
 ; APPLICANT: BARTON, Nelson
 ; APPLICANT: GARRETT, James
 ; APPLICANT: O'DONOGHUE, Eileen
 ; TITLE OF INVENTION: RECOMBINANT BACTERIAL PHYASES AND USES THEREOF
 ; FILE REFERENCE: DIVER1370-7
 ; CURRENT APPLICATION NUMBER: US/09/866,379
 ; CURRENT FILING DATE: 2001-05-24
 ; PRIOR APPLICATION NUMBER: US 09/580,515
 ; PRIOR FILING DATE: 2000-05-25
 ; PRIOR APPLICATION NUMBER: US 09/318,528
 ; PRIOR FILING DATE: 1999-05-25
 ; PRIOR APPLICATION NUMBER: US 09/291,931
 ; PRIOR FILING DATE: 1999-04-13
 ; PRIOR APPLICATION NUMBER: US 09/259,214
 ; PRIOR FILING DATE: 1999-03-01
 ; PRIOR APPLICATION NUMBER: US 08/910,798
 ; PRIOR FILING DATE: 1997-08-13
 ; NUMBER OF SEQ ID NOS: 10
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 5
 ; LENGTH: 1901
 ; TYPE: DNA
 ; ORGANISM: Escherichia coli
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)..(1901)
 ; OTHER INFORMATION: n is any nucleotide
 US-09-866-379-5

Query Match 17.0%; Score 225.2; DB 9; Length 1901;
 Best Local Similarity 53.2%; Pred. No. 1.6e-62;
 Matches 571; Conservative 0; Mismatches 489; Indels 13; Gaps 4;

QY 107 TAGAAGCTGTGGTATTTTGGTGCAGTGTGTTGGTCCGCGACCAA-ACAACGAGC 165
 DB TGGAAAGTGTGGTGAATGTTCAGTGTGATGGTGTGGTCTCCACCAAGCCAGCAAC 335
 QY 166 TTATGAATGATGTACGCGCAGATAAGTGGCCACAATGCGCGGTAAAGCGGGGTATTAA 225
 DB TGAATGAGATGTGACCCAGACGATGCGCCAACTGGCCCGGTAAACTGGTGTGACTGA 395
 QY 226 CGCCACGTTGGTGGAGTGTGTCATATGATGGGGGGTATTTATGATGATTAATTGCGCA 285
 DB CACCGCGGGTGTGAGCTAATCGCTATCTCGACATTAACCAACGCGCTCTGGTAG 455
 QY 286 GCCTTGTGTTTGTAGC---GCGGGATGTCCGCGAGAGGGGGGTATATGCAAGGAGC 342
 DB CGACCGAATTTCTGCGGAAAGAGGCTGCCCGAGTGTCTGCTCAGTTCGCGATTTGCTG 515
 DB 343 ATATGATCAACGTAACCGCTTAACCGGACAGGCAATTTCTTGTATGTTGGCTCCGGGT 402
 DB 516 ATGTGACGAGCGTACCGGTAAACAGCGCAAGCCTTTCGCGCGGGCTGGCACCTGACT 575

QY 883 ATAAAGGAGCGCATTTATTACAGCAGATAGATACGGCTTTTAACTCCCTTAACTCGATGCTC 942
Db 1050 GCGCGCCACCCCGTTATTAGATTGATCAAGACAGCGTTGACGCCCATC---CACCGC 1106
QY 943 AGGCGCAAAAGCTACCCATTTAGCCCAAAACCGGCTTTTCTCGGTGGGCATGATA 1002
Db 1107 AAAACAGGCGTATGGTGTACATTACCCACTTCACTGCTGTTTATTCGCCGACAGATA 1166
QY 1003 CCAATATTGCCAATATTGCGGGTANGCTGGAGCCGACATGCGAGCTACCGAGCAACCTG 1062
Db 1167 CTAATCTGGCAAAATCTCGCGGGGCACTGGAGCTCAACTGAGCTTCCCGGTACGCCG 1226
QY 1063 ATAATCTCCACAGGTGGGGATTGGTTTTCGAATATGCAAAATCCGGTAACCCACC 1122
Db 1227 ATACAGCGCCGACAGGTGGTGAATGTTTGAAGCTGCTGGCTGCGCTAAGCGATAACA 1286
QY 1123 AGCGCTACGTGGCGTGAAGATGTTTACCAACGATGATGATGCTGCTGCTGCTGCT 1175
Db 1287 GCCAGTGGATTGAGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1339
RESULT 10
US-09-866-379-7
; Sequence 7, Application US/09866379
; Patent No. US20020136754A1
; GENERAL INFORMATION:
; APPLICANT: DIVERGA CORPORATION
; APPLICANT: SHORT, Jay
; APPLICANT: KRETZ, Keith
; APPLICANT: GRAY, Kevin
; APPLICANT: BARTON, Nelson
; APPLICANT: GARRETT, James
; APPLICANT: O'DONOGHUE, Eileen
; TITLE OF INVENTION: RECOMBINANT BACTERIAL PHYTASES AND USES THEREOF
; FILE REFERENCE: DIVER1370-7
; CURRENT APPLICATION NUMBER: US/09/866, 379
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 09/580, 515
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 09/318, 528
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: US 09/291, 931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: US 09/259, 214
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: US 08/910, 798
; PRIOR FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 1901
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(1901)
; OTHER INFORMATION: n is any nucleotide
US-09-866-379-7
Query Match 17.0%; Score 225.2; DB 9; Length 1901;
Best Local Similarity 53.2%; Pred. No. 1.6e-62;
Matches 571; Conservative 0; Mismatches 489; Indels 13; Gaps 4;
QY 107 TAGAAGCTGGTATTATTGAGTCGCGCATGTTGCTGCGCGGACCA-ACAACGACG 165
Db 276 TGAAGAGTGGGTATTGTCAGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 335
QY 166 TTATGAATGATGTTAGCCAGATAGTGGCCCAATGGCGCGGTAAAGCGGGTATTAA 225
Db 336 TGATGAGGATGTCACCCAGACGCGATGCGCAACCTGGCGGGTAAACTGGGTGGCTGA 395
QY 226 CGCCAGCTGGTGGGAGTGGTGCATATGATGGGGGGGGTTTATTGATGATTAATTCTT 285

Db 396 CACCGCGNGTGGTGAAGCTAATCGCTATCTCGGACATTTACCAACGCCAGCGTCTGGTAG 455
QY 286 GCCTGGTGTGTTAGCG--GCGGATGTCGGGAGAGGGGGGTATATCACAGGCGAG 342
Db 456 CCGACGATTTCTGGCGAAAAGGGCTGCCCGCAGTCTGGTCAGGTCCGATTTATGCTG 515
QY 343 ATATCGATCAACGTACCCGCTTAAACCGGACAGGCAATTTCTTTGATGGTGGTGGCGGT 402
Db 516 ATGTCGACGAGCTACCGGTAAACAGGCGAAGCTTTGCGCGCGGGGTGGCACTGACT 575
QY 403 GTGGTTTACCGCTGATATATCAGCCGATTTGAAAAGACCGATCCCTCTGTCATCCGG 462
Db 576 GTGCAATTAACGTATACCCAGGACAGATAGCTCAGTCCCGATCCGTTATTTATCTC 635
QY 463 TAGACATGGCGGTGTAAAGTTAGCAACGACCAACAGATAAAGCGATTTGAAGAACAT 522
Db 636 TAAAACTGGCGTTTCCCACTGGATAACGCGAACGTCAGTACGCGATCTCAGCAGGG 695
QY 523 TGGGCGGCGCTTAGATACGGTGGAGCCGCTAGCTAAACCTTTTCCGAGATGGGG 582
Db 696 CAGGAGGTCAATTGCTGACTTTACCGGCGCATCGGCAACCGCGCTTTGCGGAACTGGAAC 755
QY 583 ACCTGCTGAATTTTGGCGCTTCTCTTTATTGCAAAATCTTTGCAACAGCAAGAAAACGT 642
Db 756 GGGTCTTTAATTTCCGCAATCAAACTTTGTCCTTAAACGTGAGAAAACAGGACGAAAGCT 815
QY 643 GTGATTTGGCCACTTTGGCGCCAAATGAAGTTAAGCTTAAATAAGAGAGGACAAAAGTGA 702
Db 816 GTTCATTAACGAGCGCATTTACATCGGAATCAAGGTGAGCGCGA-----CAATGTCT 869
QY 703 CCCTCAGTGGGCCACTGGCGTTATCATCGACATTTGGGTGAAATCTTCTTATTACAAAACG 762
Db 870 CATTAAACGGTGGTGAAGCTTCGATCAATGCTGACGGAGATATTTCTCTGCAACAG 929
QY 763 CACAAGCCATGCCAGAGGTTGCTGGCAACGCTTAAAGGGGCGGAGAAATGGGTATCTCT 822
Db 930 CACAGGGAATGCCGAGCGCGGGTGGGGAAGGATCACCGATTACACCCAGTGGAAACCT 989
QY 823 TATTGTCAATTACATAACGCGCAATTTAATTTAATGGCAAAAACACCGCTATATGCCCGTC 882
Db 990 TGCTAAGTTTGCATAACGCGCAATTTTATTGCTACACGCGCAGAGGTTGCCCGCA 1049
QY 883 ATAAAGGAGCGCATTTATTACAGCAGATAGATACGGCTTTTAAACCTTTCAACTGATGCTC 942
Db 1050 GCCGCGCACCCCGTTATTAGATTGATCAAGACAGCGTTTGAACGCCCCCATC---CACCGC 1106
QY 943 AGGCGCAAAAGCTACCCATTTACGCCCAAAACCGGGCTTTGTTCTCGGTGGGCATGATA 1002
Db 1107 AAAACAGGCGTATGGTGTGACATTACCCACTTCACTGAGCTCAACTGAGCGCTTCCCGGT 1166
QY 1003 CCAATATTGCAATATTGCGGGTANGCTGGAGCCGACATGCGAGCTACCGAGCAACCTG 1062
Db 1167 CTAATCTGGCAAAATCTCGCGGGCGCACTGGAGCTCAACTGAGCGCTTCCCGGTGAGCGCG 1226
QY 1063 ATAATCTCCACAGGTGGGGATTGGTTTTCGAATATGCAAAATCCGGTAACCCACC 1122
Db 1227 ATACAGCGCGCGAGGTGGTGAATGTTTGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1286
QY 1123 AGCGCTACGTGGCGTGAAGATGTTTACCAACGATGATGATGCTGCTGCTGCTGCTGCT 1175
Db 1287 GCCAGTGGATTGAGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1339

RESULT 11
US-10-156-660-3
; Sequence 3, Application US/10156660
; Publication No. US20030103958A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay M.
; APPLICANT: Kretz, Keith
; APPLICANT: Gray, Kevin A.
; APPLICANT: Barton, Nelson R.


```

; APPLICANT: Garrett, James B.
; APPLICANT: O'Donoghue, Eileen
; APPLICANT: Mathur, Eric J.
; TITLE OF INVENTION: PHASES, NUCLEIC ACIDS ENCODING THEM
; TITLE OF INVENTION: AND METHODS FOR MAKING AND USING THEM
; FILE REFERENCE: 09010-029007
; CURRENT APPLICATION NUMBER: US/10/156,660
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: US 09/866,379
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 09/580,515
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 09/318,528
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: US 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: US 09/259,214
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: US 08/910,798
; PRIOR FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1901
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (188)...(1483)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 403
; OTHER INFORMATION: n = A,T,C or G
US-10-156-660-3

```

```

Query Match      17.0%; Score 225.2; DB 15; Length 1901;
Best Local Similarity 53.2%; Pred. NO. 1.6e-62;
Matches 571; Conservative 0; Mismatches 489; Indels 13; Gaps 4;

```

```

QY 107 TAGAACGTGTGTTATTTTGGTCCCATGTTTGGTCCGCGACCAA-ACAACGCGC 165
DB 276 TCGAAAGTGTGTTATTTGTCAGTCTGTCATGTTGTCGTCCTCCACCAAGCCAGCGAAC 335
QY 166 TTATGAATGATTTACGCGAGTATAGTCCGCGATGCGCGGTAAGCGGGGTTATTA 225
DB 336 TGATGAGGATGTACCCCGCAGCGATGCGCAACCTGCGCGGTAAGCGGGTGGTGA 395
QY 226 CGCCACGTGTCGCGAGTGTGTCATGATGCGGGGGGTTTATGTTGATTACTTCGCA 285
DB 396 CACCGCGGTGTGAGCTAATCGCTTATCTCGGACATTCACACCGCGGTCTGTGTAG 455
QY 286 GCCTTGGTTTGTAGCG---GCGGATGTCGCGCAGAGCGGGGGTATATGACAGGCGAG 342
DB 456 CGACCGGATTTGTCGCGAAAGAGGCTGCGCGAGTCTGTGTCAGGTGCGGATTTTGTG 515
QY 343 ATATCGATCAACGTACCGCTTAACCGGACAGCGATTTCTTGATGTTGGTCTCGGGGT 402
DB 516 ATGTGACGAGGTACCGGTAAACAGCGAAGCGCTTCGCGCGGGGTGCGACCTGACT 575
QY 403 GTGGTTTACCGTGTATATCAGCGCGATTTGAAAGACCGATCCCTGTTCCATCCGG 462
DB 576 GTGCAATAACCGTACATACCCAGCGAGATGTCAGTCCGATCCGTTATTTAATCTC 635
QY 463 TAGAGACTGGTGTGTAGTTAGTACAGCGCAACAGATTAAGCGATTGAAGAACGAT 522
DB 636 TAAAGAACTGGCGTTTGGCCAACTGGATAACGCGAAGCTGACTGACCGGATCTCTCAGAGG 695
QY 523 TGGCGGGGCGGTAGATACGGTGTGACGCGGTACGCTTAAACCTTTTCCAGATGGGG 582
DB 696 CAGGAGGTCAATGCTGACTTACCGGGCATCGCAACCGCGTTTCGCACTGGAAC 755
QY 583 ACGTCTGAATTTGGCGGCTTCTCCTTATGTAATTTTGCACAGCAAGAAACGT 642

```

```

756 GGGTGTCTTAATTTTCCGCAATCAAACTGTGCTTAAACGTGAGAAACAGGACGAAAGCT 815
643 GTGATTTTCCACATTTGGCGCAATGAAGTTAAAGTTAAATAAGAGGACAAAGTGA 702
816 GTTCAATTAAGCAGGAGGATTAACATGGAATCTAGGTGAGGCCGA-----CAATGCT 869
703 CCCTCAGTGGCCACATGGCGTTATCATCGACATTTGGTGAATCTTTTATTTATCAAAAG 762
870 CATTAAACCGGTGCGGTAAAGCTTCGATCAATGCTGACGAGATATTTCTCTCTCAACAG 929
763 CACAGCCATGCGCAGAGGTTGCTGCGACGCTAAAGGGGCGGAGATTTGGGTATCT 822
930 CACAGGATGCGCGAGCGGGTGGGAAGATACCGATTCAACAGTGGAAACACT 989
823 TATTGTCTATTACATACGCGCAATTTAATTTAATGCAAAAAACCGGTATATCGCCGTC 882
990 TGCTAAGTTTGCATACGCGCAATTTAATTTGCTAACACGCGCAGAGGTTGCGCGCA 1049
883 ATAAAGGAGCGCATTTATACAGCAGATAGATACGCTTTAACCTTTCACTGGATGCTC 942
1050 GCGCGCCACCCCGTTATTAGATTGATCAAGACAGCGTTGACGCCCATC---CACCGC 1106
943 AGGGGCAAAAGCTTACCCATTTTCAGCCCAAAACCGGTCTTTGTTCTCGTGGGCATGATA 1002
1107 AAAACAGCGGTATGTTGATGACATTACCCACATTCAGTGTCTTTATCGCGGACGATA 1166
1003 CCAATATTGCGAATATTGCGGTATGCTGGAGCGGACTGGCAGCTACCCGAGCAACTG 1062
1167 CTAATCTGCGCAATCTCGCGCGCAGCTGAGCTCAACTGAGCTTCCCGTCAAGCGG 1226
1063 ATAATACTCCACCGGTGGGATTTGTTTGAATATGGAATAATCGGATAAACCC 1122
1227 ATAACACGCGCGCAGGTGTGAACTGGTGTGTTGAACGCTGGCGTAAAGCGATAACA 1286
1123 AGGCTACGTTGCGGTGAAGATGTTCTACCAACGATGATCAGTTGGTAAAT 1175
1287 GCCAGTGGATTGAGTTTCGCTTTCGCTTTCAGACTTTACAGCAGATGCGTGTAT 1339

```

RESULT 12

```

US-10-282-122A-7167
; Sequence 7167, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625

```


PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 7167
LENGTH: 1299
TYPE: DNA
ORGANISM: Escherichia coli
US-10-282-122A-7167

Query Match
Best Local Similarity 53.2%; Score 224.6; DB 13; Length 1299;
Matches 571; Conservative 0; Mismatches 489; Indels 13; Gaps 4;

QY 107 TAGAAGTGTGGTATTTCAGTCGCCATGCTGCTCGCCGACCA-ACAAGCGCAGC 165
Db 89 TGGAAAGTGTGGTATTTCAGTCGCCATGCTGCTCGCCGACCA-ACAAGCGCAGC 148
QY 166 TTATGAATGATGTACGCCAGATAGTGGCCACAATGGCCGGTAAAGCGGGGTATTTAA 225
Db 149 TGATGCAGGATGTCACCCAGACGATGGCCAACTGGCCGGTAAAGCTGGGTGCTGA 208
QY 226 CGCACGTGTGGGAGTTGGTCAATGATGGGGGGTTTATGTGATTACTTTTCGCA 285
Db 209 CACCGCGGTGTGAGCTAATCGCTATCTCGCAATACCAACGCCAGCTGTGGTAG 268
QY 286 GCCTTGTGTTTGTAGG--GCGGGATGTCGCCAGAGGGGGGTATATGCACGGCAG 342
Db 269 CCGACGATTTGCTGGGAAAAGGGTGGCCGAGTCTGTGAGTGGCGATTATTTGCTG 328
QY 343 ATATCGATCAAGTACCGCTTAAACGGACAGAGATTTCTGTGTTGGTGGTCCGGGT 402
Db 329 ATGTCGACGAGCTACCCGTAAACAGCGAAGCCCTGCGCCGGGTGGCACCTGACT 388
QY 403 GTGTTTTCAGCTGATTAATCAGCGGATTTGAAAAGACCGATCCCTGTTCCATCCG 462
Db 389 GTGATTAACGGTACATACCGGACAGATGCTCCAGTCCCGATCGTTATTTAATCTC 448
QY 463 TAGAGACTGGCGTGTGTAAGTTAGACAAACGCAAAAGATTAAGGATTAAGAACGAT 522
Db 449 TAAAACTGGGTTTGGCAACTGGATAACCGAAGCTGACTGACGGGATCTCAGCAGGG 508
QY 523 TGGCGGGCGTTAGATACGGTACGGTACGCTACGCTAAACCTTTTCCCGACATGGGG 582
Db 509 CAGGAGGTCAATGCTGACTTTACCGGGCATCGGCAACCGGCTTTCGGAACTGAAAC 568
QY 583 ACGTGCTGAATTTTGGCGCTTCTCTTATTGCAAAATCTTTGCAACGAAAGGAAAAACGT 642
Db 569 GGGTGTCTAATTTTCCGCAATCAAACTTGTGCTTAAACGCTGAGAAACAGGACGAAAGCT 628
QY 643 GTGATTTTGGCCATTTTGGCGCCAAATGAAGTTAAGTTAATAAGAGGAGCAAAAGTGA 702
Db 629 GTTCATTAAACGAGGCAATTAACATCGGAATCAAGGTGAGCGCGA-----CAATGTCT 682
QY 703 CCCTCAGTGGGCCACTGGCGTTATCATCGACATTTGGGTGAAATCTTCTTATTACAAAACG 762
Db 683 CATTAACCGGTGGCGTAAGCTCCATCAATGCTGACGGAGATTTTCTCTCAACAG 742
QY 763 CAAAGCCATGCGAGGTTCCCTGGCAACGGCTAAAGGGGGGAGAAATGGGTATCTT 822
Db 743 CACAGGGAATGCCGGAGCGGGTGGGGAAGGATCACCGATTACACAGTGGAAACACCT 802
QY 823 TATTGTCAATACATAACGCGCAATTTAATTTAATGGCAAAAACACCGTATATCGCCGCTC 882
Db 803 TGCTAAGTTTGCATTAACGGCGAATTTTATTTGCTTACAAACGCAAGAGGTTGCCGCA 862
QY 883 ATAAAGGAGCGCCATTATACAGCAGATAGATACGGCTTTAAACCTTTCAACTGATGCTC 942

Db 863 GCOCGCCACCCCGTTATTAGATTTCAGATCAAGCAGGTTGACGCCCATC---CACCGC 919
QY 943 AGGGGCAAAAGCTACCAATTTAGCCCAAAACCGGGTCTTTTCTCGTGGGATGATA 1002
Db 920 AAAAACAGGGGTATGTTGATGACATTACCCACTTCAGTGTCTTTATCGCCGACACGATA 979
QY 1003 CCATATTTCCCAATATTGCGGGTATGCTGGAGCCGCTGCGAGTACCCGAGCAACCTG 1062
Db 980 CTAATCTGGCAATCTCGCGCGCACTGGAGCTCAACTGAGCTTCCCGTTCAGCGG 1039
QY 1063 ATAATATCTCCACAGGTGGGGGATTTGTTTGAACCTATGCAAAATTCGGATTAACACC 1122
Db 1040 ATAACACGCCGCAAGTGGTGAACCTGTTTGAACGCTGGCGTGGCTAAGCGATAACA 1099
QY 1123 AGCCTAGTTGCGGTGAAGATGTTTACCAAAAGATGATCAGTTCGGTAAT 1175
Db 1100 GCCAGTGAATTCAGGTTTCGCTGCTTCCAGACTTTTACAGCAGATCGGTGAT 1152

RESULT 13
US-09-866-379-9
Sequence 9, Application US/09866379
Patent No. US20020136754A1
GENERAL INFORMATION:
APPLICANT: DIVERSA CORPORATION
APPLICANT: SHORT, Jay
APPLICANT: KRETZ, Keith
APPLICANT: GRAY, Kevin
APPLICANT: BARTON, Nelson
APPLICANT: GARRETT, James
APPLICANT: O'DONOGHUE, Eileen
TITLE OF INVENTION: RECOMBINANT BACTERIAL PHYTASES AND USES THEREOF
FILE REFERENCE: DIVER1370-7
CURRENT APPLICATION NUMBER: US/09/866,379
CURRENT FILING DATE: 2001-05-24
PRIOR APPLICATION NUMBER: US 09/580,515
PRIOR FILING DATE: 2000-05-25
PRIOR APPLICATION NUMBER: US 09/318,528
PRIOR FILING DATE: 1999-05-25
PRIOR APPLICATION NUMBER: US 09/291,931
PRIOR FILING DATE: 1999-04-13
PRIOR APPLICATION NUMBER: US 09/259,214
PRIOR FILING DATE: 1999-03-01
PRIOR APPLICATION NUMBER: US 08/910,798
PRIOR FILING DATE: 1997-08-13
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn version 3.1
SEQ ID NO 9
LENGTH: 1901
TYPE: DNA
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)..(1901)
OTHER INFORMATION: n is any nucleotide
US-09-866-379-9

Query Match
Best Local Similarity 16.9%; Score 223.6; DB 9; Length 1901;
Matches 570; Conservative 0; Mismatches 490; Indels 13; Gaps 4;

QY 107 TAGAAGTGTGGTATTTCAGTCGCCATGCTGCTCGCCGACCA-ACAAGCGCAGC 165
Db 276 TGGAAAGTGTGGTATTTCAGTCGCCATGCTGCTCGCCGACCA-ACAAGCGCAGC 335
QY 166 TTATGAATGATGTACGCCAGATAGTGGCCACAATGGCCGGTAAAGCGGGGTATTTAA 225
Db 336 TGATGCAGGATGTCACCCAGACGATGGCCAACTGGCCGGTAAAGCTGGGTGCTGA 395
QY 226 CGCACGTGTGGGAGTTGGTCAATGATGGGGGGTTCATGATGGGGGGTTCATGATTTTTCGCA 285
Db 396 CACCGCGGTGGTGGTGAAGCTATCTCGGACATTCACCAACGCCAGGCTCTGGTAG 455

286 GCCTTGGTTGTTAGCG---CCGGATGTCGGCAGAGGGGGGGTATATGACAGGCAG 342
Db CCGACGGAATTCCTCGCGAAAAGAGCGCTGCCGCGAGTCTGTGTCAGTCGCGAATATTGCTG 515
Qy ATATCGATCAACGTAACCGCTTAACCGGACAGAGCAATTCCTGTATGTTGGTCTCCGGGGT 402
Db ATGTCGACGAGCGTACCGTAAACAGCGAAGCCTTCGCGCGCGGCTGGCACTGACT 575
Qy GTGGTTGACCGTGCATTAATCAGCGCGAATTTGAAAAGACCGATCCCTCTTCCATCCGG 462
Db GTGCAATAACCGGTACATACCCAGGAGATACGTCCAGTCCGATCCGTTATTTAATCCTC 635
Qy TAGAGACTGGCGTGTGAAGTTAGACACGACAAACAGATTAAGCGGATTTGAAGAAGCAT 522
Db TAAAAACTGGCGTTTGCCAACTGGATACCGGAACGCTGACTGACGCGATCCTCAGCAGGG 695
Qy TGGGGGGCGCTTAGATAGCGTAGCGAGCAGCGCTAGCTAAACCTTTTGGCCAGATGGGG 582
Db CAGGAGGCTCAATTGCTGACTTTACCGGGCATCGGCAACGGCGTTTTCGGGAATGGAAC 755
Qy ACGTGCTCAATTTTGGCGCTTCTCGTTATTCGCAATCTTTGCAACAGCAGGAAGAAACGT 642
Db GGGTGTCTTAATTTTCGCAATCAAACTTGTGCTTAAACGTGAGAAACAGGACGAAGCT 815
Qy GTGATTTTGGCCACTTTTGGCGGCAATGAAGTTAAGCTTAATAAAGAGAGGACAAAAGTGA 702
Db GTTCAATTAACGAGGCAATTAACCATCGGAATCAAGGTGAGCGCGCA-----CAATGCT 869
Qy CCTCAGTGGCGCACTGCGCTTATCATCGACATTTGGGTGAATCTTCTATTACAAAACG 762
Db CATTAACCGTGGCTAGCTCGATCATGCTGACGAGATATTCTCTCTGCAACAG 929
Qy CACAAGCCATGCGAGAGTTCCTGGCAACGGCTTAAAGGGGGCGGAGATTTGGGTATCCT 822
Db CACAGGGAATGCGGAGCGGGGTGGGAAGGATCACCGATTACACCGGTGAACACCT 989
Qy TATTCATTAACATTAACGCGCAATTTAATTTAATGGCAAAACACCGTATATCCCGCTC 882
Db TGCTAAGTTTGCATTAACGCGCAATTTTGTGCTTACACGACGCGCAGAGTTGCGCGCA 1049
Qy ATAAAGGACGCGCAATTTATACAGCAGATAGATACGGCTTTAAACCTTCAACTGATGCTC 942
Db GCCGCGCACCGCTTATTAGATTTGATCAAGACAGCGCTTACGCGCCCATC---CACCGC 1106
Qy AGGGCCAAAGCTACCAATTTACGCCCCAAACCGGCTTCTGCTCGTGGTGATGATA 1002
Db AAAAAACAGCGCTATGGTGTGACATTACCCACTTTCAGTCTGTTTATCCCGGACAGATA 1166
Qy CCAATATTGCCAATATTGGGGTATGCTGGAGCGGACTGGCAGCTACCCGAGCAACCTG 1062
Db CTAATCTGCAAACTCTCGCGCGCAGCTGGAGCTCAACTGGACGCTTCCCGGTACGCGG 1226
Qy ATAATACTCCACAGTGGGGGATTTGTTTGAACCTATGCAAAATCGGATACCAACC 1122
Db ATAACAGCGCGGAGTGGTGAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1286
Qy AGCGCTTACGTTGGGTGAAGATGTTCTACCAACAGATGGATCAGTTGGGTAAAT 1175
Db GCCAGTGGATTGAGTTTTCGCTGGCTTCCAGACTTTACAGCAGATGGGTAT 1339

RESULT 14
US-09-777-566A-1
; Sequence 1, Application US/09777566A
; Patent No. US20010055789A1
; GENERAL INFORMATION:
; APPLICANT: DIVERSA CORPORATION
; APPLICANT: SHORT, Jay
; APPLICANT: KREIZ, Keith
; TITLE OF INVENTION: RECOMBINANT BACTERIAL PHYTASES AND USES THEREOF
; FILE REFERENCE: DIVER1370-6
; CURRENT APPLICATION NUMBER: US/09/777,566A

; CURRENT FILING DATE: 2001-06-11
; PRIOR APPLICATION NUMBER: US 09/318,528
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: US 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: US 09/259,214
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: US 08/910,798
; PRIOR FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatenIn version 3.0
; SEQ ID NO 1
; LENGTH: 1323
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1320)
; NAME/KEY: misc_feature
; LOCATION: (1)..(1323)
; OTHER INFORMATION: n is any nucleotide
US-09-777-566A-1

Query Match 16.5%; Score 218.8; DB 9; Length 1323;
Best Local Similarity 52.8%; Pred. No. 1.6e-60;
Matches 567; Conservative 0; Mismatches 493; Indels 13; Gaps 4;

Qy 107 TAGAAGTGTGTTATTTTGTAGTCCCATGTTGCTGCGCGACCAA-ACAAGCGAGC 165
Db 89 TGAAGAAGTGTGTTATTTGTCAGTCTGTCATGTTGCTGCTCCAAACCAAGGCGACGCAAC 148
Qy 166 TTATGATCATGTTAGTCCGCGACAGATAAGTGCACCAATGCGCGTAAAGCGGGGTATTTAA 225
Db 149 TGATGAGATGTTTACCCCGACGATGCGCACTGCGCGGTAAATCTGGTGGTCTCA 208
Qy 226 CGCCACGTCGTGCGAGTTGGTCAATTGATGGGGGGTTTTATGGTGAATATCTTCGCA 285
Db 209 CACCGCGGTGCTGAGCTAATCGCTATCTCGGACATTTACCAAGCGCCAGCGCTCTGGTAG 268
Qy 286 GCCTTGTGTTAGCG---GCGGATGTCGCGAGAGGGGGGTATATGACAGGCAG 342
Db 269 CGACGAGTGTGCGGAAAAAGGCTGCCGAGTCTGTCAGTCTGTCAGTCTGCGATTTTGTG 328
Qy 343 ATATCGATCAAGCTACCCGCTTTAAACCGGACAGGCATTTCTTGTATGGTGTGGTCCGGGT 402
Db 329 ATGTGACGAGCTACCCGTAACAGGCGAAGCCTTCGCCCGCGGCTGGCACTGACT 388
Qy 403 GTGGTTGACCGTGCATTAATCAGCGCGATTTGAAAAGACCGATCCCTGTTCATCCCG 462
Db 389 GTCAATAACCGGTACATACCCAGGAGATACGTCAGTCCGATCCGATCGTTATTTAATCCTC 448
Qy 463 TAGAGACTGCGGTGTAAAGTTAGACACGACCAACAGATAAAGCGATTTGAAGACGAT 522
Db 449 TAAAACTGGCGTTTGCCAACTGGATAACGCGAAGCTGACTGACGCGATCCTCAGCAGG 508
Qy 523 TGGCGCGCGGTAGATAGGTGAGCGGCTACGCTAAACCTTTTGGCCAGATGGGG 582
Db 509 CAGGAGGGTCAATTGCTGACTTTTACCGGCGATCGGCAACCGGCTTCGCGAATCGAAC 568
Qy 583 ACCTGTGTAATTTTGGGGTCTTCTTATTTGCAAAATCTTTTGCAACAGAGGAAAAACGT 642
Db 569 GGGTGTAAATTTTCGCAATCAAACTGTCCTTAAACGTGAGAAACAGGACGAAGCT 628
Qy 643 GTGATTTGCCCACTTTTGGCGCAATGAAGTTAACTTAATTAAGAGGACGCAAAAGTGA 702
Db 629 GTTCAATTAACGAGGCAATTAACATCGGAACTCAAGGTGAGCGCGCA-----CAATGTCT 682
Qy 703 CCCTCAGTGGGCCACTGGGGTTATCATCGACATTTGGGTGAATCTTCTTTATTACAAAACG 762
Db 683 CATTAACCGGTGGGTAAAGCTCGCATCAATGCTGACGAGATATTTCCTCTGCAACAG 742
Qy 763 CACAAGCCATGCGAGGTTGCTCGCAACGGCTTAAAGGGCGGGAATTTGGGTATCCT 822

Db 743 CACAGGGAATGCGGAGCGGGGTGGGGAAGGATCACCGATTACACCAAGTGAACACCT 802
 Qy 823 TATTGTCAATACATAACGGCAATTTAAATTAATGCAAAAAACACCGTATATCGCCGTC 882
 Db 803 TGCTAAGTTTGCAATACGGCAATTTTATTGCTACAAACGACGCGAGAGTTGCGCGCA 862
 Qy 883 ATAAAGGAGCGCAATTTATACAGCAGATAGATACGGCTTTAAACCTTTCACTGAGTGC 942
 Db 863 GCGCGCCACCGCGTTATTGGATTGATATGCGCAGCTTGAACCGCCATCCAC---GCG 919
 Qy 943 AGGGGCAAAAGCTTACCCATTTCAGCCCAAAACCGGGTCTTGTTCCTCGGTGGCATGATA 1002
 Db 920 AAAAAACAGCGATGCTGTGACATTACCCACTTCAGTACTGTTTATTGCGGACAGATA 979
 Qy 1003 CCAATATGCGAATATTGCGGTATGCTGGAGCCGACTGGCAGCTACCGAGCAACCTG 1062
 Db 980 CTAATCTGCAATCTCGCGCGCACTGGAGCTCAACTGGAGCTTCCCGTCAACCGG 1039
 Qy 1063 ATAATACTCCACAGGTGGGGATTGTTTTTGAACATAGGCAAAATCCGATACCCACC 1122
 Db 1040 ATAAACAGCGCGCAGGTGTGAACCTGTTTGAACGCTGGCGTGGCTTAACCGATAACA 1099
 Qy 1123 AGCGTACGTTGGGTGAAGTGTCTACCAACGATGGATCAGTTGCGTAAAT 1175
 Db 1100 GCCAGTGGATTGAGTTTCGCTGCTTCCAGACTTTACAGCAGATGCGTAT 1152

RESULT 15

US-09-866-379-1
 ; Sequence 1, Application US/09866379
 ; Patent No. US200201367541
 ; GENERAL INFORMATION:
 ; APPLICANT: DIVERSA CORPORATION
 ; APPLICANT: SHORT, Jay
 ; APPLICANT: KREY, Keith
 ; APPLICANT: GRAY, Kevin
 ; APPLICANT: BARTON, Nelson
 ; APPLICANT: GARRETT, James
 ; APPLICANT: O'DONOGHUE, Eileen
 ; TITLE OF INVENTION: RECOMBINANT BACTERIAL PHYTASES AND USES THEREOF
 ; FILE REFERENCE: DIVER1370-7
 ; CURRENT APPLICATION NUMBER: US/09/866,379
 ; CURRENT FILING DATE: 2001-05-24
 ; PRIOR APPLICATION NUMBER: US 09/580,515
 ; PRIOR FILING DATE: 2000-05-25
 ; PRIOR APPLICATION NUMBER: US 09/318,528
 ; PRIOR FILING DATE: 1999-05-25
 ; PRIOR APPLICATION NUMBER: US 09/291,931
 ; PRIOR FILING DATE: 1999-04-13
 ; PRIOR APPLICATION NUMBER: US 09/259,214
 ; PRIOR FILING DATE: 1999-03-01
 ; PRIOR APPLICATION NUMBER: US 08/910,798
 ; PRIOR FILING DATE: 1997-08-13
 ; NUMBER OF SEQ ID NOS: 10
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 1
 ; LENGTH: 1323
 ; TYPE: DNA
 ; ORGANISM: Escherichia coli
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (1)..(1323)
 ; OTHER INFORMATION: n is any nucleotide
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(1323)
 ; OTHER INFORMATION:
 US-09-866-379-1

Query Match 16.5%; Score 218.8; DB 9; Length 1323;
 Best Local Similarity 52.8%; Pred. No. 1.6e-60;
 Matches 567; Conservative 0; Mismatches 493; Indels 13; Gaps 4;
 Qy 107 TAGAAGCGTGGTATTATTGAGTCGCCCATGGTGTTCGCTGCCGACCAA-ACAACGCAGC 165

Db 89 TGGAAAGTGTGGTGAATTTGTCAGTCGTCTATGGTGTGCGTGTCCAAACCAAGGCCACGCAAC 148
 Qy 166 TTATGAATGATGTTACCGCAGATAGTGGCCACATGCGCGGTAAAGACGGGGTATTATAA 225
 Db 149 TGATGCAAGATGTCAACCCACAGACGATGGCCAAACCTGCGCGGTAAACCTGGTGTGGCTGA 208
 Qy 226 CGCCACGTGGTGGCGAGTTGGTCAATTGATGGGGGGTTTTATGCGATTTACTTTTCGCA 285
 Db 209 CACCGCGNGGTGGTGAAGTATCGCTATCTCGGACATTACCAACGCCAGCGTCTGGTAG 268
 Qy 286 GCCTTGGTTTGTAGCG---GCGGGATGTCGGCAGAGGGGGGGGTATATATGCAAGCGAG 342
 Db 269 CCGACGATGCTGCGCGAAAGGGGCTGCCCGAGTCTGGTCAAGTCTGGTCTCGATTTATTCTG 328
 Qy 343 ATATCGATCAACGTACCGCGTTAAACCGGACAGGCAATTTCTTTGATGGTGTGGTCCCGGGT 402
 Db 329 ATGTCGACGAGCGTACCGGTAAACACGCGAGCGCTTCGCGCGCGGGTCCGACTGACT 388
 Qy 403 GTGTTTTGACCGTGCATTAATCAGCCGATTTGAAAAAGACCGATCCCTCTTCATCCGG 462
 Db 389 GTGCAATAACCGTACATACCCAGCAGATACGTCAGTCCCGATCCGTTATTTAATCCTC 448
 Qy 463 TAGAGACTGGCGTGTGAAGTTAGAACACGACAAACAGATAAAGCGATTGAAGAACGAT 522
 Db 449 TAAAACTGCGGTTTGCCAACTGGATTAACGCGAACGTGACTGACCGATCCTCAGCAGGG 508
 Qy 523 TGGCGCGCGCTTAGATACCGGTGAGCGCGCTACGCTAAACCTTTTCCCGAGATGGGG 582
 Db 509 CAGGAGGCTCAATTTGCTGACTTTTACCGGGCATCGGAAACCGGGCTTTCGCGAACTGGAAC 568
 Qy 583 ACGTGCTGAATTTTGGCGCTTCTCTTATTGCAAACTTTTGCACACGACGAGGAAACGCT 642
 Db 569 GGGTGTCTTAATTTTCCCAATCAAACTTGTGCTTAAACGCGGAGAACAGACGAAAGCT 628
 Qy 643 GTGATTTTGCCTTTTGGCGCCAATGAAGTTAAAGTTAAAGAGGAGGACAAAAGTGA 702
 Db 629 GTTCATTAAACGACGAGCATTTACCATCGAACTCAAGGTGAGCGCGCA-----CAATGTCT 682
 Qy 703 CCCTCAGTGGCGCACTGGCGTTATCTCGACATTTGGTGAATCTTCTTATTATCAAAACG 762
 Db 683 CATTAAACCGCGGTGAAGCTCGCATCAATGCTGACGGAGATATTCTCTCTGCAACAAG 742
 Qy 763 CACAAGCCATGCCAGAGGTTGCCGTGCAACGCGCTAAAGGGGCGGAGAAATTTGGGTATCCT 822
 Db 743 CACAGGGAATGCCGAGCGGGGTGGGAAAGGATCACCGATTCAACACGAGTGAACACCT 802
 Qy 823 TATTGCTATTACATAACGCGCAATTTAATTAATGCAAAACACCGTATATCGCGGTC 882
 Db 803 TGCTAAGTTTGCATAACGCGCAATTTTATTGCTACAAACGCGCAGAGGTTGCCCGCA 862
 Qy 883 ATAAAGGAGCGCCATTTATTACAGCAGATAGATACGGCTTTAAACCTTTCAACTGGATGCTC 942
 Db 863 GCGCGCCACCGCTTATTGGATTGATCATGCGAGGGTTGACGCCCATCCAC---GCG 919
 Qy 943 AGGGGCAAAAGCTACCCATTTACGCCCAAAACCGGGTCTTGTTCCTCGGTGGCGATGATA 1002
 Db 920 AAAAAACAGCGGTATGGTGTGACATTACCCACTTCAGTACTGTTTATTGCGCGACAGATA 979
 Qy 1003 CCAATATTGCCAATATTGCGGTATGCTGGAGCCGACTGGCAGCTACCGGACCAACCTG 1062
 Db 980 CTAATCTGGCAATCTCGCGCGCGCACTGGAGCTCACTGAGCGCTTCCCGTCAAGCGG 1039
 Qy 1063 ATAATACTCCACAGGTGGGGATTGTTTTTGAACATATGGCAAAATCCGGATAAACACC 1122
 Db 1040 ATAAACGCGCGCAGGTGGTGAACCTGTTTGAACGCTGGCGTCCGCTAAAGCGATAACA 1099
 Qy 1123 AGCGTACGTTGGGTGAAGTGTCTACCAACGATGGATCAGTTGCGTAAAT 1175
 Db 1100 GCCAGTGGATTGAGTTTCGCTGCTTCCAGACTTTACAGCAGATGCGGTAT 1152

Search completed: May 3, 2004, 07:14:51

Job time : 572.19 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2004, 22:18:08 ; Search time 2501.04 Seconds
(without alignments)
15820.390 Million cell updates/sec

Title: US-10-021-723B-3
Perfect score: 1325
Sequence: 1 agctgatttagataatgc.....aacctgcctgcacatctaa 1325

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pin:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_pbg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	115.2	8.7	529	13	BU713770
c 2	107.4	8.1	616	12	BU074127
3	103.4	7.8	746	13	CA093060
c 4	90.6	6.8	595	14	CF326092

C	5	75.4	5.7	393	10	BE520240
	6	63.4	4.8	531	13	BU713769
	7	62.6	4.7	383	9	AK036132
c	8	62.6	4.7	383	9	AK036134
	9	50.4	3.8	885	13	EX425603
	10	49.2	3.7	707	12	BU619443
c	11	48.8	3.7	208	12	BM409016
c	12	48.8	3.7	208	12	BM412806
c	13	46.2	3.5	354	9	AA545747
c	14	44	3.3	1201	13	BU381961
c	15	43.2	3.3	1101	29	CNS00KX2
c	16	41.8	3.2	520	13	EX477052
c	17	40.8	3.1	858	13	BU930958
c	18	40.4	3.0	179	10	BE461872
c	19	40.2	3.0	553	9	AU280373
c	20	40.2	3.0	594	13	EX473143
c	21	40.2	3.0	602	13	EX479427
c	22	40.2	3.0	626	9	AL701247
c	23	40.2	3.0	643	10	BE314719
c	24	40.2	3.0	649	10	BE313446
c	25	40.2	3.0	688	10	BE298415
c	26	40.2	3.0	692	13	EX482609
c	27	40.2	3.0	722	9	AL709367
c	28	40.2	3.0	893	13	BU931710
c	29	40.2	3.0	895	10	BE796373
c	30	40.2	3.0	899	13	EX386580
c	31	40.2	3.0	954	13	BU845888
c	32	40.2	3.0	969	13	EX333117
c	33	40.2	3.0	989	10	BE314590
c	34	40.2	3.0	1064	12	BM912235
c	35	40.2	3.0	1067	13	EX397055
c	36	40.2	3.0	1126	12	BM462187
c	37	40.2	3.0	1139	12	BM916904
c	38	40.2	3.0	1171	13	EX394513
c	39	40.2	3.0	1201	9	AL529590
c	40	40.2	3.0	1201	13	EX354373
c	41	40.2	3.0	1705	12	BM456811
c	42	39.8	3.0	1201	13	EX385091
c	43	39.6	3.0	1190	13	EX377026
c	44	39.4	3.0	723	13	EX727085
c	45	39.4	3.0	992	13	EX394359

ALIGNMENTS

RESULT 1
BU713770
LOCUS

DEFINITION

BU713770 529 bp mRNA linear EST 23-OCT-2003
SJAABUG02 Adult sJc 7/94 Schistosoma japonicum cDNA similar to
pdb|1DKPIA Chain A, Crystal Structure Of Phytate Complex Of
Escherichia Coli Phytase At Ph 6.6. Phytate Is Bound With Its
3-Phosphate In The Active Site. Hg2+ Cation Acts As An
Intermolecular Bridge, mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BU713770 GI:28321126
EST.
Schistosoma japonicum
Schistosoma japonicum

REFERENCE
AUTHORS

Hu, W., Yan, Q., Shen, D. K., Liu, F., Zhu, Z. D., Song, H. D., Xu, X. R.,
Wang, Z. J., Rong, Y. P., Zeng, L. C., Wu, J., Zhang, X., Wang, J. J.,
Xu, X. N., Wang, S. Y., Fu, G., Zhang, X. L., Wang, Z. Q., Brindley, P. J.,
McManus, D. P., Xue, C. L., Feng, Z., Chen, Z., and Han, Z. G.
Evolutionary and biomedical implications of a Schistosoma japonicum
complementary DNA resource
Nat. Genet. 35 (2), 139-147 (2003)

TITLE
JOURNAL
MEDLINE
COMMENT

Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai

201203, P. R. China
Tel: 86-21-50801919 (ex. 45)
Fax: 86-21-50801922
Email: hanzg@chc.sh.cn.
Location/Qualifiers
1. 529
/organism="Schistosoma japonicum"
/mol_type="mRNA"
/strain="Chinese (Anhui) strain"
/db_xref="taxon:6182"
/sex="Male and female"
/tissue_type="Whole body"
/dev_stage="Adult worms"
/lab_host="Mouse and rabbit"
/clone_lib="Adult SJC 7/94"
/note="Vector: Lambda ZAP-III XR.; Site 1: EcoR I; Site 2:
XhoI I; Several hundred adult Schistosoma japonicum
(Anhui, P.R. China, strain), of mixed sex, were perfused
from the mesenteries of experimentally infected mice and
rabbits at the Queensland Institute of Medical Research,
Brisbane, Australia (QIMR), and stored for several months
in liquid nitrogen. Subsequently, mRNA was isolated at the
QIMR from lysates of these worms by oligo dT
chromatography, using a kit from Pharmacia. The mRNA was
then shipped to Clontech, Palo Alto, CA, USA, who
constructed a cDNA library. First strand synthesis was
primed with an oligo-dT-XhoI primer and synthesized using
M-MuV reverse transcriptase. Second strand and synthesis was
accomplished with RNase H and T4 DNA polymerase. The
double stranded cDNA was ligated to EcoRI linkers,
digested with EcoRI and XhoI, and ligated into the
phagemid vector lambda ZAP II XR. After construction of
this directional library by Clontech, it was returned to
the QIMR. During analysis of the library at the QIMR, we
have found that a small percentage, 2% to 3%, of the
clones contain inserts that appear to be highly homologous
to sequences from salmonid fishes, as determined by
homology comparisons using BLAST and by Southern
hybridization analysis to genomic DNA from salmon (Sigma
Chemical Co., St. Louis, MO) under stringent washing
conditions. The remainder of the clones appear to contain
S. japonicum sequences."

FEATURES

source

Db 364 CAGCCGGATACACGCCGCCAGGTGGTGAATGTTTGAACGCTGCGCTAAGC 423
QY 1116 AACCAACAGCCGCTACGTTGCGGTGAAGATGTTTACCAACAGTGGATCTGCGTAAT 1175
Db 424 GATAACAGCCAGTGGATTGAGTTTGGCTGCTTCCAGACTTTTACAGCAGATGCGTAT 483
RESULT 2
BJ074127/c 616 bp mRNA linear EST 29-SEP-2003
LOCUS BJ074127 NIBB Mochii normalized Xenopus tailbud library Xenopus
DEFINITION laevis cDNA clone XL090j06 5', mRNA sequence.
ACCESSION BJ074127
VERSION BJ074127.1 GI:17504316
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Xenopus laevis (Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 616)
REFERENCE Kitayama, A., Terasaka, C., Mochii, M., Ueno, N., Shin-i, T. and
Kohara, Y.
Expressed genes in X. laevis embryo
Unpublished (2001)
Contact: Tadashi Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tsun@genes.nig.ac.jp
The information of this clone is available through the following
URL.
http://xenopus.nibb.ac.jp.

FEATURES

source

1. 616
Location/Qualifiers
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="XL090j06"
/tissue_type="whole embryo"
/dev_stage="stage 25"
/clone_lib="NIBB Mochii normalized Xenopus tailbud
library"

ORIGIN

Query Match 8.1%; Score 107.4; DB 12; Length 616;
Best Local Similarity 50.1%; Pred. No. 1.8e-22;
Matches 298; Conservative 0; Mismatches 291; Indels 6; Gaps 1;
QY 305 GGGATGTCGGCAGAGGGGGGGTATATGCACAGCAGATATCGATCAACGTACCGCTT 364
Db 589 GGGTGTCCCGCAGTCTGGTCAGGTCGCGATTATTCTGATGTCGACGAGCGTACCCGTAA 530
QY 365 AACCGCAGAGGCAATTTCTTTGATGGTGTGGCTCCGGGGTGTGGTTTGACCGTGCATAATCA 424
Db 529 AACGCGGAAGCTTCGCGCGCGGCTGGCACCCTGACTGTGCAATAACCGTACATACCCA 470
QY 425 GGCGGATTTGAAAAGACCGATCCCTGTTCCATCCGTAGAGACTGGGTTGTAGTT 484
Db 469 GGCAGATACGTCAGTCCCGATCCGTTATTAACTCTCTAAAACTGGGTTTGGCACT 410
QY 485 AGACACGACACAAACAGATAAAGCGATTGAAGACGATTGGCGGGCCGTTAGATACGCT 544
Db 409 GGNATACGCGAAGCTGACTGACGCGATCCTCAGAGGGCAGGAGGTCAATTGTGACTT 350
QY 545 GAGCCAGCGCTACGCTAAACCTTTTGGCCAGATGGGGACGCTGCTGAATTTTGGGCTTC 604
Db 349 TACCGGCACTCGCGCAAAACGCGCTTTCCGCACTGGAAACGGGTGCTTAATTTTCCCAATC 290
QY 605 TCCTTATTGCAATCTTTTCCACAGCAGAAAAACGTGTGATTTTGGCCCACTTTGGCGSC 664

```

Db      289 AAACTTGTGCTTAAACGTCGAGAAACAGGACGAAAGCTGTTCAATTAACGAGGCATTACC 230
QY      665 CAATGAAGTTAACTTAATAAGAGGACAGAAAGTGACCTCAGTGGGCGCACTGGCGTT 724
Db      229 ATCGGAACTCAGGTGAGCCCGA-----CAATGCTCATTAACCGGTGCGGTAGCGCT 176
QY      725 ATCATGCACATTCGGTGAAATCTTCTTATTACAAAACGCAAGCCATCCAGAGTTGC 784
Db      175 CGCATCAATCTGACGGAGATATTTCTCTGCAACAAAGCACTGGGAATGCCGAGCCGGG 116
QY      785 CTGCGACGCTAAAGGGGCGGAGATGGGTATCTTATTCATTAATCATTAACGAGCGCA 844
Db      115 GTGGGAGGATCACCGATTACACCACTGGAACACCTTGCTAAGTTTGCATTAACGCGCA 56
QY      845 ATTTAATTTAATGGCAAAACACCGTATATCGCCCGTCATAAAGGACGCCATT 899
Db      55 ATTTTATTGCTACAACGCGCAGAGGTTGCCCGCAGCGCGCCACCCCGTTA 1

```

RESULT 3
CA093060
LOCUS
DEFINITION
3', mRNA sequence.
CA093060.1 GI:34946367
EST.
Saccharum officinarum
Saccharum officinarum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Saccharum.
1 (bases 1 to 746)
Vettore, A.L., da Silva, F.R., Kemper, E.L. and Arruda, P.
The libraries that made SUCEST
Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
Contact: Arruda P
Centro de Biologia Molecular e Engenharia Genetica
Universidade Estadual de Campinas
Caixa Postal 6010, 13083-970, Campinas SP, Brazil
Tel: 55 19 3788 1137
Fax: 55 19 3788 1089
Email: parruda@unicamp.br
Clone distribution: clone distribution information can be found
through the Brazilian Clone Collection Center (BCCC) at
http://www.bcccenter.fcav.unesp.br
Plate: 001 row, D column: 10
Seq primer: M13/Forward primer.
Location/Qualifiers
1. 746
/organism="Saccharum officinarum"
/mol_type="mRNA"
/db_xref="taxon:4547"
/clone="SCCCCL2001D10"
/lab_host="XlBlue MRF"
/clone_lib="CL2"
/note="Organ: Pool of sugarcane calli submitted to low
(40C) and high (37 C) temperature stress; Vector:
pBluscript; Site 1: EcoRI; Site 2: XhoI; An
unidirectional cDNA library generated from [Pool of
sugarcane calli submitted to low (40C) and high (37 C)
temperature stress]. cDNA was prepared from polyA+ mRNA
using ZAP - cDNA Synthesis Kit (Stratagene). The
double-strand cDNAs were fractionated in a sepharose CL-2B
40cm-columns and fragments sizing between 0.8 and 1.5 Kb
were directionally cloned into the vector. Details of
each source of RNA and library construction can be
obtained at http://sucest.lad.ic.unicamp.br/public"

Query Match
Best Local Similarity 7.8%; Score 103.4; DB 13; Length 746;
53.5%; Pred. No. 3.8e-21;

```

Matches 257; Conservative 0; Mismatches 219; Indels 4; Gaps 2;
QY      696 AAAGTGACCCCTCAGTGGGCGCACTGGCGTTATCATCGCATTTGGGTGAAATCTTCTTATTA 755
Db      265 AATGCTCATTAACCGGTGCGGTAAAGCTCGCATCAATGCTGACGGAGATATTTCTCTG 324
QY      756 CAAAAGCACAAGCCATGCCAGAGTTCCCTGSCAACGGCTAAAGGGGCGGGAATGG 815
Db      325 CAACAAGCACAGGAATGCCGAGCCCGGGTGGGGAAGGATCACCGATTACACACAGTGG 384
QY      816 GTATCTTATTGTCTATTAACATGAACGCGCAATTTAATTTAATGGCAAAACACCGTATATC 875
Db      385 AACACCTTGCTAAGTTTGCATAACGCGCAATTTTATTTGCTACACGCGCAGAGGTT 444
QY      876 GCGCGTATAAAGGAGCGCCATTTATACAGCAGATAGATACGGCTTTAACCCCTTCAACTG 935
Db      445 GCGCGCAGCCCGCCACCCCGTTATTAGATTGATCAAGACAGCGTTGACGCCCATC-- 502
QY      936 GATGCTCAGGCGCAAAAGCTACCCATTTCAGCCCAAAACCGGGTCTTGTTCCTCGGTGG 995
Db      503 -CACCGCAAAACAGCGGTATGGTGTGACATCCACTTCAGTGTCTTTATCGCGGA 561
QY      996 CATGATACCAATATTGCCAATATTTCGCGGTATGTCGGAGCCGACTGCGACGTACCCGAG 1055
Db      562 CACGATATAATCTGGCAATCTCGGGCGGCACCTGGAGCTCAACTTGGACGCTTCCCGGT 621
QY      1056 CAACCTGATATATCTCCACAGGTGGGGATTGTTTTTGAACATATGCAAAATCCGGAT 1115
Db      622 CAGCCGATAACACGCCCGCAGGTGGTGAATGG-TGTTGAACGCTGCGCTGCTGCTAAGC 680
QY      1116 AACCACCAAGCGCTAGTTGCGGTGAAGATGTTTACCAACAGATGGATCAGTGGGTAAAT 1175
Db      681 GATAACACCAAGTGGATTTCAGGTTTCGCTGCTCTCCAGACTTTACAGCANATGCGTGT 740

```

RESULT 4
CF326092/c
LOCUS
DEFINITION
595 bp mRNA linear EST 18-AUG-2003
JMT1--05-B09-g1 AtJMT-overexpressing transgenic rice lambda phage
cDNA library (JMT1) Oryza sativa cDNA clone JMT1--05-B09, mRNA
sequence.
CF326092
EST.
CF326092.1 GI:33800445
Oryza sativa
Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzae; Oryza.
1 (bases 1 to 595)
Kim, J.S., Jun, K.M., Cheong, P.J., Kim, M.J., Lee, T.H., Shin, Y.C.,
Song, S.I., Kim, J.K., Kim, Y.-K. and Nahm, B.H.
Large-scale Sequencing Analysis of Rice ESTs
Unpublished (2003)
Contact: Nahm B.H.
Genomics and Genetics Institute, GreenGene Biotech Inc.; Division
of Bioscience and Bioinformatics, Myongji University
Yongin, Gyeonggi, Korea
Tel: 82 31 330 6193
Fax: 82 31 321 6355
Email: bhnah@gbio.com, bhnah@bio.myongji.ac.kr.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
1. 595
/organism="Oryza sativa"
/mol_type="mRNA"
/cultivar="Nackdong"
/db_xref="taxon:4530"
/clone="JMT1--05-B09"
/tissue_type="leaf"
/dev_stage="14 days after germination"
/lab_host="E.coli SOLR"
/clone_lib="AtJMT-overexpressing transgenic rice lambda
phage cDNA library (JMT1)"

FEATURES
source
1. 746
/organism="Saccharum officinarum"
/mol_type="mRNA"
/db_xref="taxon:4547"
/clone="SCCCCL2001D10"
/lab_host="XlBlue MRF"
/clone_lib="CL2"
/note="Organ: Pool of sugarcane calli submitted to low
(40C) and high (37 C) temperature stress; Vector:
pBluscript; Site 1: EcoRI; Site 2: XhoI; An
unidirectional cDNA library generated from [Pool of
sugarcane calli submitted to low (40C) and high (37 C)
temperature stress]. cDNA was prepared from polyA+ mRNA
using ZAP - cDNA Synthesis Kit (Stratagene). The
double-strand cDNAs were fractionated in a sepharose CL-2B
40cm-columns and fragments sizing between 0.8 and 1.5 Kb
were directionally cloned into the vector. Details of
each source of RNA and library construction can be
obtained at http://sucest.lad.ic.unicamp.br/public"

Query Match
Best Local Similarity 7.8%; Score 103.4; DB 13; Length 746;
53.5%; Pred. No. 3.8e-21;

ORIGIN
Query Match
Best Local Similarity 7.8%; Score 103.4; DB 13; Length 746;
53.5%; Pred. No. 3.8e-21;

/note="Vector: pBluescript SK(+); Site 1: EcoRI; Site 2: XhoI; cDNA was inserted into lambda Uni-ZAP XR vector at 5' end with EcoRI and 3' end with XhoI site. mRNA was prepared from Arabidopsis Ussuinolate Carboxyl methyltransferase overexpression line."

ORIGIN

Query Match 6.8%; Score 90.6; DB 14; Length 595;
Best Local Similarity 55.9%; Pred. No. 4.3e-17;
Matches 193; Conservative 0; Mismatches 149; Indels 3; Gaps 1;

QY 831 TTACATACGCCCAATTAATTAATGCGAAMACCGTATATGCGCGTCATAAAGG 890
Db 592 TTGCATACGCCCAATTAATTAATGCTACACGCGCAGAGTTGCGCGCGCGCC 533
QY 891 AGCCATTATTACAGCAGATAGATACGGCTTTAAACCTTTAACTGATGCTCAGGGCAA 950
Db 532 ACCCGTTATTAGATTGATCAAGACAGGTTGACGCCCATCCAC---CCGAAACAG 476
QY 951 AAGCTACCCATTTCAGCCCAACCGGGTCTTTTCCTCGGTGGCATGATACCAATTT 1010
Db 475 GCGTAGGTGTGACATTACCCACTTCAGTGTCTGTTTATCGCGGACAGATACTAATCTG 416
QY 1011 GCCAATATTGCGGTATGTCGGAGCGCACTGGCACTACCGAGCACTGATAATACT 1070
Db 415 GCAATCTCGGCGGCACTGAGCTCACTGGAGCTTCCGTCAGCGCATACAG 356
QY 1071 CCACAGGTGGGGATTTGTTTTTGAATATGCGAAATCGGATPACACACGCGTAC 1130
Db 355 CGCCAGGTGGTGAATCTGTTTGAACGCTGGCTTAAGCGATAACAGCCAGTGG 296
QY 1131 GTTCGGGTGAAGATGTTCTACCAACGATGATGATGCGTAAT 1175
Db 295 ATTGAGTTTCTGCTGCTTCCAGACTTTACAGCAGATGCGTGAT 251

RESULT 5
BES20240/c
LOCUS M1B12STM Arabidopsis developing seed Arabidopsis thaliana cDNA
DEFINITION clone M1B12 5', mRNA sequence.
ACCESSION BES20240.1 GI:9778242
VERSION BES20240
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 393)
White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de Ilarduya, O., Jaworski, J.G., Ohlrogge, J. and Benning, C.
A new set of Arabidopsis expressed sequence tags from developing seeds. The metabolic pathway from carbohydrates to seed oil
Plant Physiol. 124 (4), 1582-1594 (2000)
JOURNAL 20567808
MEDLINE 11115876
PUBMED
COMMENT Contact: Benning, C
Dept. of Biochemistry & Molecular Biology
Michigan State University
224 Biochemistry, Michigan State University, East Lansing, MI 48824, USA
Tel: 517 355 1609
Fax: 517 353 9334
Email: benning@msu.edu
Michigan State University DNA Sequencing Facility Arabidopsis Biological Resource Center, The Ohio State University, 309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX: 6142920603 TEL: 614292371.

FEATURES
source
1. .393
/organism="Arabidopsis thaliana"
/mol_type="mRNA"

/strain="Columbia"
/db_xref="taxon:3702"
/clone="M1B12"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/clone_lib="Arabidopsis developing seed"
/notes="Organ: Developing seed; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoII"

ORIGIN

Query Match 5.7%; Score 75.4; DB 10; Length 393;
Best Local Similarity 51.0%; Pred. No. 2.6e-12;
Matches 175; Conservative 0; Mismatches 168; Indels 0; Gaps 0;

QY 339 GCAGATATGATCAACGTACCCGCTTAACGCGACAGCATTTCTTATGATGGTGGCTCG 398
Db 387 GCTGATGTCGACGAGGTTACCCGTAACAGCGGAGCTTTGCGCGCGGNTGGACCT 328
QY 399 GGGTGTGGTTTACCGTCATATCAAGCGCGATTTGAAAAAGACCGATCCCTGTTCAT 458
Db 327 GACTGTGCAATACCGTACATACCCAGGCGAGTACGTCAGTCCCGATCCGTTATTAA 268
QY 459 CCGGTAGAGACTGGCGTGTAGTTAGTAAACGCAACAGATAAAGCGATTGAAGAA 518
Db 267 CCTCTAAAACCTGGCGTTTCCCACTGATACGGAACGTAAGTCACTGACGCACTCAGC 208
QY 519 CGATCGGCGGCGGTAGATACGTTGACGCGCTACGCTAAACCTTTTGGCCAGATG 578
Db 207 AGGCGAGGAGGGTCAATTGCTGACTTTACCGGCGCATCGGCAACGCGCTTCCGCACTG 148
QY 579 GGGGAGCTGCTGAATTTTGGGCTTCTCTTATTGCAATCTTTGCAACAGCAAGGAAA 638
Db 147 GAACGGGTGCTTAATTTTCCGCAATCAAACTTGCTTAAACGTCGAGAAACAGGACGA 88
QY 639 ACGTGTGATTTTGGCCCACTTTGCGGCAATGAAGTTAACGTTA 681
Db 87 AGCTGTTTCAATACGAGCATTTACCATCGGAACCTCAGGTGA 45

RESULT 6
BU713769
LOCUS BU713769.1 GI:28321125
DEFINITION 531 bp mRNA linear EST 23-OCT-2003
SJAABUG01 Adult SJC 7/94 Schistosoma japonicum cDNA similar to SP|P07102|PPA_ECOLI_PERIPLASMIC_APPA_PROTEIN_PRECURSOR [INCLUDES: PHOSPHONHYDRIE PHOSPHOHYDROLASE (PH 2.5 ACID PHOSPHATASE) (AP); 6-PHYTASE 1, mRNA sequence.
ACCESSION BU713769
VERSION BU713769.1
KEYWORDS EST.
SOURCE Schistosoma japonicum
ORGANISM Schistosoma japonicum
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidae; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE 1 (bases 1 to 531)
Hu, W., Fan, Q., Shen, D.K., Liu, F., Zhu, Z.D., Song, H.D., Xu, X.R., Wang, Z.J., Rong, Y.P., Zeng, L.C., Wu, J., Zhang, X., Wang, J.J., Xu, X.N., Wang, S.Y., Fu, G., Zhang, X.L., Wang, Z.Q., Brindley, P.J., McManus, D.P., Xue, C.L., Feng, Z., Chen, Z. and Han, Z.G.
Evolutionary and biomedical implications of a Schistosoma japonicum complementary DNA resource
Nat. Genet. 35 (2), 139-147 (2003)
JOURNAL 22879925
MEDLINE
COMMENT Contact: Zegu Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China
Tel: 86-21-50801919 (ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn.
Location/Qualifiers
1. .531
/organism="Schistosoma japonicum"

/mol_type="mRNA"
/strain="Chinese (Anhui) strain"
/db_xref="taxon:6182"
/sex="Male and female"
/tissue_type="Whole body"
/dev_stage="Adult worms"
/lab_host="Mouse and rabbit"
/clone_lib="Adult SJC 7/94"
/note="Vector: Lambda ZAP-III XR.; Site 1: EcoRI; Site 2: XhoI I; Several hundred adult Schistosoma japonicum (Anhui, P.R. China, strain), of mixed sex, were perfused from the mesenteries of experimentally infected mice and rabbits at the Queensland Institute of Medical Research, Brisbane, Australia (QIMR), and stored for several months in liquid nitrogen. Subsequently, mRNA was isolated at the QIMR from lysates of these worms by oligo dT chromatography, using a kit from Pharmacia. The mRNA was then shipped to Clontech, Palo Alto, CA, USA, who constructed a cDNA library. First strand synthesis was primed with an oligo-dT-XhoI-primer and synthesized using M-MuLV reverse transcriptase. Second strand synthesis was accomplished with RNase H and T4 DNA polymerase. The double stranded cDNA was ligated to EcoRI linkers, digested with EcoRI and XhoI, and ligated into the phagemid vector lambda ZAP II XR. After construction of this directional library by Clontech, it was returned to the QIMR. During analysis of the library at the QIMR, we have found that a small percentage, 2% to 3%, of the clones contain inserts that appear to be highly homologous to sequences from salmonid fishes, as determined by homology comparisons using BLAST and by Southern hybridization analysis to genomic DNA from salmon (Sigma Chemical Co., St. Louis, MO) under stringent washing conditions. The remainder of the clones appear to contain S. japonicum sequences."

ORIGIN

Query Match 4.8%; Score 63.4; DB 13; Length 531;
Best Local Similarity 49.1%; Pred. No. 2.2e-08;
Matches 196; Conservative 0; Mismatches 201; Indels 2; Gaps 1;
QY 696 AAGTACCTCAGTGGGCCACTGCGGTATCATCGACATTTGGTGAATCTTCTATTATTA 755
Db 7 AATGTCATTAACCGGTGCGGTAAAGCTCGCATCAATGCTGACGAGATGGGCTCCTG 66
QY 756 CAAACGACACAGCCATCGAGGTTGCCGCAACGCTTAAAGGGCGGAGATGG 815
Db 67 CAACAGCACAGGGAATCCGGAGCCGGGTGGGAAGGATCACGATTCACACAGTGG 126
QY 816 GTATCCTTATTGCTATTACATAACGCGCAATTTAATTTAATGGCAAAACACCGTATATC 875
Db 127 AACACCTTGTGAAGTTGCTAATACGCGCAATTTTATTTGCTACACGCAACGAGGTT 186
QY 876 GCCGCTCATAAAGGACGCTTATTAACGAGATAGATACGGCTTTAACCTTCACTG 935
Db 187 GCCCGACGCGGCCACCGCGTATTAGATTGATCAAGACAGCGTGTACGCG - CCGGTT 244
QY 936 GATGCTCAGGGGCAAAAGCTACCCATTTTCAGCCCAAAACCGGGTCTTGTCTCGGTGGG 995
Db 245 TATCATTAATGATTTCATGTGGAGAGACAATTCCTCTTGTAGTACGAAAGGCGGC 304
QY 996 CATGATACCAATATTGCCMAATTTGGGGTATGCTGGGAGCGGACTGCGACCTACCGAG 1055
Db 305 CACGAGCCTGAGCTGATAAACTCGCGCGGCGACTGAGGCTCAACTGAGCGCTGTCCAGG 364
QY 1056 CAACCTGATAATCTCCACAGGTGGGATTTGTTT 1094
Db 365 CAGCCCGATACACGCCCGCAGGAGCGGAATGGATGTT 403

RESULT 7
AW036132
LOCUS
AW036132 383 bp mRNA linear EST 18-MAY-2001

DEFINITION EST274508 tomato seed, TAMU Lycopersicon esculentum cDNA clone cLBE1E23 similar to periplasmic phosphoanhydride phosphohydrolase precursor, putative, mRNA sequence.
AW036132
VERSION AW036132.1 GI:5894811
KEYWORDS EST.
SOURCE Lycopersicon esculentum (tomato)
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon. 1 (bases 1 to 383)
Alcala, J., Vrebalov, J., White, R., Matern, A.L., Vision, T., Holt, I.E., Liang, F., Upton, J., Renning, C.M., Craven, M.B., Fujii, C.Y., Bowman, C.L., Nierman, W., Fraser, C.M., Venter, J.C., Martin, G.B., Tanksley, S.D. and Giovannoni, J.
Generation of ESTs from tomato seed tissue
Unpublished (1999)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
3 Prime sequence.
Location/Qualifiers
1. 383
/organism="Lycopersicon esculentum"
/mol_type="mRNA"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLBE1E23"
/tissue_type="seeds"
/dev_stage="quiescent seed"
/lab_host="XLI-Blue MRF"
/clone_lib="tomato seed, TAMU"
/notes="Vector: pBlueScript SK(-); Site 1: EcoRI; Site 2: XhoI; cLBE - Tomato Seed EST Library. Directionally cloned cDNAs inserted into pBlueScript SK(-) at 5' end with EcoRI and 3' end with XhoI site."

ORIGIN

Query Match 4.7%; Score 62.6; DB 9; Length 383;
Best Local Similarity 49.8%; Pred. No. 3.2e-08;
Matches 159; Conservative 0; Mismatches 159; Indels 0; Gaps 0;
QY 365 AACCGACAGGCAATTTCTGTGTTGGTTCGGGGTGTGTTTACCGTGCATATCA 424
Db 4 ALCAGGGAAGCCTTCGCGCGGCTGCACTGTGCAATTAACCGTACATACCA 63
QY 425 GCGCGATTTGAAAAGACCGATCCCTGTTCCATCGGTAGAGACTGGCGTGTGAAGTT 484
Db 64 GCGAGATACGTCAGTCCGATCCGTTATTTATCTCTTAAACTGGCGTTGCCACT 123
QY 485 AGACAACGCAACAAAGATAAGCGATTGAAGAACGATTGGGGCGGCCGTTAGATACGGT 544
Db 124 GGATAACGGAACGATGACTGACGCGATCTCAGCAGGCGAGGAGGTCAATTGCTGACTT 183
QY 545 GAGCAGCGGTACGCTAAACCTTTTCCCGAGATGGGGACGTGCTCAATTTTCGCGCTTC 604
Db 184 TACCGGCGATCGCAACCGCGTTTCGCGACTGGAACCGGTCTTAATTTCCGCAATC 243
QY 605 TCCTTATTGCAATCTTTTGCACAGCAAGAAACAAACGTTGATTTTGGCCACTTTCCGGC 664
Db 244 AAACCTTGTGCTTAAACGTAAGAAACAGGACGAAAGCTGTTCATTAAACGAGCATACC 303
QY 665 CANTGAAGTAACGTTA 681
Db 304 ATCGAAGTCAAGGTGA 320

RESULT 8
AW036134/c
LOCUS
AW036134 383 bp mRNA linear EST 18-MAY-2001

<p>RESULT 10</p> <p>LOCUS</p> <p>DEFINITION</p> <p>ACCESSION</p> <p>VERSION</p> <p>KEYWORDS</p> <p>SOURCE</p> <p>ORGANISM</p> <p>REFERENCE</p> <p>AUTHORS</p> <p>TITLE</p> <p>JOURNAL</p> <p>COMMENT</p>	<p>BJ619443</p> <p>BJ619443</p> <p>707 bp mRNA linear EST 01-OCT-2003</p> <p>Xenopus laevis cdna clone XL189Q24 5', mRNA sequence.</p> <p>BJ619443</p> <p>BJ619443.1 GI:37258203</p> <p>Xenopus laevis (African clawed frog)</p> <p>Xenopus laevis</p> <p>Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus.</p> <p>1 (bases 1 to 707)</p> <p>Kitayama,A., Terakasa,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara,Y.</p> <p>Expressed genes in X. laevis embryo</p> <p>Unpublished (2001)</p> <p>Contact: Tadasu Shin-i</p> <p>Center for Genetic Resource Information</p> <p>National Institute of Genetics</p> <p>1111 Yata, Mishima, Shizuoka 411-8540, Japan</p> <p>Tel: 81-559-81-6856</p> <p>Fax: 81-559-81-6855</p> <p>Email: tshini@genes.nig.ac.jp</p> <p>The information of this clone is available through the following URL.</p> <p>http://xenopus.nibb.ac.jp.</p>	<p>FEATURES</p> <p>source</p> <p>Location/Qualifiers</p> <p>1..707</p> <p>/organism="Xenopus laevis"</p> <p>/mol_type="mRNA"</p> <p>/db_xref="taxon:8355"</p> <p>/clone="XL189Q24"</p> <p>/tissue_type="whole embryo"</p> <p>/dev_stage="stage 10.5"</p> <p>/clone_lib="NIBB Mochii normalized Xenopus early gastrula library"</p>	<p>ORIGIN</p> <p>Query Match 3.7%; Score 49.2; DB 12; Length 707;</p> <p>Best Local Similarity 52.4%; Pred. No. 0.00091;</p> <p>Matches 133; Conservative 0; Mismatches 118; Indels 3; Gaps 1;</p> <p>QY 771 ATCCGAGGTTCCCTGGCAACGCTAAAGGGCGGAGATTGGGTATTCCTATTGTCA 830</p> <p>Db 2 ATCCGGAGCGGGGTGGGAAGGATCCCGATTACACGATGGAACACCTTGTCTAAGT 61</p> <p>QY 831 TTACATAACCGCAATTTAATTTAATGGCAAAACACCGTATATCCCGTCAATGAAGG 890</p> <p>Db 62 TTGCATAACCGCAATTTAATTTGCTACTAACGCAACGAGGTTCCCGCGAGCGCGC 121</p> <p>QY 891 ACCGCTATTATACAGCAGATAGATACGCTTTAACCCTTCACTGATGCTCAGGGGCA 950</p> <p>Db 122 ACCCGTTATTAGATTGATCAAGACAGCGTTGACGCCCATC---CAGCGCAAAACAG 178</p> <p>QY 951 AAGCTACCCATTTTCAAGCCAAACACCGGCTCTTCTCGGTGGGCATGATACCAATATT 1010</p> <p>Db 179 GCGTATGGTGTGACATTACCCACTTCTAGTGTCTTTATCGCCGACACGATCTAATCTG 238</p> <p>QY 1011 GCCAATATTCGGG 1024</p> <p>Db 239 GCAATCTCGGG 252</p>	<p>RESULT 11</p> <p>LOCUS</p> <p>DEFINITION</p> <p>ACCESSION</p> <p>VERSION</p> <p>KEYWORDS</p> <p>SOURCE</p> <p>ORGANISM</p> <p>REFERENCE</p> <p>AUTHORS</p> <p>TITLE</p> <p>JOURNAL</p> <p>COMMENT</p>	<p>BM409016</p> <p>BM409016</p> <p>208 bp mRNA linear EST 22-JAN-2002</p> <p>tomato breaker fruit Lycopersicon esculentum cDNA clone</p> <p>BM409016</p> <p>BM409016.1 GI:18260646</p> <p>EST.</p> <p>Lycopersicon esculentum (tomato)</p> <p>Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.</p> <p>1 (bases 1 to 208)</p> <p>Alcala,J., Vrebalov,J., White,R., Vision,T., Karamycheva,S.A., Tsai,J., Bougri,O., Kirkness,E., Utterback,T., Van Aken,S.,</p>	<p>FEATURES</p> <p>source</p> <p>Location/Qualifiers</p> <p>1..208</p> <p>/organism="Lycopersicon esculentum"</p> <p>/mol_type="mRNA"</p> <p>/cultivar="TA496"</p> <p>/db_xref="taxon:4081"</p> <p>/clone="cLEG46H24"</p> <p>/tissue_type="pericarp"</p> <p>/dev_stage="breaker"</p> <p>/lab_host="SOLR"</p> <p>/clone_lib="tomato breaker fruit"</p> <p>/note="Vector: pBluescriptSKMCluadapt; Site 1: EcoRI; Site 2: XhoI; supplier: Boyce Thompson Institute; sequencing: The Institute for Genomic Research. Fruit were harvested at the breaker stage (first sign of lycopene accumulation on the blossom end of fruit). Fruit were cut in half and the seeds and locules were discarded prior to freezing the pericarp."</p>	<p>ORIGIN</p> <p>Query Match 3.7%; Score 48.8; DB 12; Length 208;</p> <p>Best Local Similarity 57.1%; Pred. No. 0.00062;</p> <p>Matches 89; Conservative 0; Mismatches 67; Indels 0; Gaps 0;</p> <p>QY 305 GGGATGTCGCGACAGGGGGGTATATGACAGGAGATATCGATCAAGTACCCGTT 364</p> <p>Db 159 GGGTGGCCCGCAGTCTGGTCAGGTCGCGATTATTTGCTGACAGGCGTACCCGTAA 100</p> <p>QY 365 AACCGGACAGGCAATTTCTTGATGTGTGGTCCGGGTGGTGTGACCGTGCAATCA 424</p> <p>Db 99 AACGGCGAAGCTTCGCGCGCGGCTGCGACCTGACTGTGCAATACCGTACATACCA 40</p> <p>QY 425 GCGCGATTGAAAAGACCGATCCCTGTTCATCC 460</p> <p>Db 39 GCGAGATACGTCCTCAGTCCCGATCCGTTATTTAATCC 4</p>	<p>RESULT 12</p> <p>LOCUS</p> <p>DEFINITION</p> <p>ACCESSION</p> <p>VERSION</p> <p>KEYWORDS</p> <p>SOURCE</p> <p>ORGANISM</p> <p>REFERENCE</p> <p>AUTHORS</p> <p>TITLE</p> <p>JOURNAL</p> <p>COMMENT</p>	<p>BM412806</p> <p>BM412806</p> <p>208 bp mRNA linear EST 22-JAN-2002</p> <p>tomato breaker fruit Lycopersicon esculentum cDNA clone</p> <p>BM412806</p> <p>BM412806.1 GI:18264436</p> <p>EST.</p> <p>Lycopersicon esculentum (tomato)</p> <p>Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.</p> <p>1 (bases 1 to 208)</p> <p>Alcala,J., Vrebalov,J., White,R., Vision,T., Karamycheva,S.A., Tsai,J., Bougri,O., Kirkness,E., Utterback,T., Van Aken,S.,</p>	<p>FEATURES</p> <p>source</p> <p>Location/Qualifiers</p> <p>1..208</p> <p>/organism="Lycopersicon esculentum"</p> <p>/mol_type="mRNA"</p> <p>/cultivar="TA496"</p> <p>/db_xref="taxon:4081"</p> <p>/clone="cLEG46H24"</p> <p>/tissue_type="pericarp"</p> <p>/dev_stage="breaker"</p> <p>/lab_host="SOLR"</p> <p>/clone_lib="tomato breaker fruit"</p> <p>/note="Vector: pBluescriptSKMCluadapt; Site 1: EcoRI; Site 2: XhoI; supplier: Boyce Thompson Institute; sequencing: The Institute for Genomic Research. Fruit were harvested at the breaker stage (first sign of lycopene accumulation on the blossom end of fruit). Fruit were cut in half and the seeds and locules were discarded prior to freezing the pericarp."</p>	<p>ORIGIN</p> <p>Query Match 3.7%; Score 48.8; DB 12; Length 208;</p> <p>Best Local Similarity 57.1%; Pred. No. 0.00062;</p> <p>Matches 89; Conservative 0; Mismatches 67; Indels 0; Gaps 0;</p> <p>QY 305 GGGATGTCGCGACAGGGGGGTATATGACAGGAGATATCGATCAAGTACCCGTT 364</p> <p>Db 159 GGGTGGCCCGCAGTCTGGTCAGGTCGCGATTATTTGCTGACAGGCGTACCCGTAA 100</p> <p>QY 365 AACCGGACAGGCAATTTCTTGATGTGTGGTCCGGGTGGTGTGACCGTGCAATCA 424</p> <p>Db 99 AACGGCGAAGCTTCGCGCGCGGCTGCGACCTGACTGTGCAATACCGTACATACCA 40</p> <p>QY 425 GCGCGATTGAAAAGACCGATCCCTGTTCATCC 460</p> <p>Db 39 GCGAGATACGTCCTCAGTCCCGATCCGTTATTTAATCC 4</p>
--	--	--	---	--	---	--	--	--	---	--	--

Roaming, C.M., Fraser, C.M., Martin, G.B., Tanksley, S.D. and Giovannoni, J.
Generation of ESTs from tomato fruit tissue, breaker stage (2002)
Unpublished (2002)
Contact: CUGI

Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
This clone is available through the Clemson University Genomics Institute

Seq primer: T3
Location/Qualifiers
1..208
/organism="Lycopersicon esculentum"
/mol_type="mRNA"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLSG61013"
/tissue_type="pericarp"
/dev_stage="breaker"
/lab_host="SOLR"
/clone_lib="tomato breaker fruit"
/note="Vector: pBluescriptSKmCUadapt; Site 1: EcoRI; Site 2: XhoI; supplier: Boyce Thompson Institute; sequencing: The Institute for Genomic Research. Fruit were harvested at the breaker stage (first sign of lycopene accumulation on the blossom end of fruit). Fruit were cut in half and the seeds and locules were discarded prior to freezing the pericarp."

ORIGIN

Query Match 3.7%; Score 48.8; DB 12; Length 208;
Best Local Similarity 57.1%; Pred. No. 0.0062;
Matches 89; Conservative 0; Mismatches 67; Indels 0; Gaps 0;
QY 305 GGGATGTCGGCAGAGGGGGGTATATGACAGGAGATATCATCAAGTACCCGCTT 364
Db 159 GGGCTGCCCGAGCTGCTGAGGTCGCGATTATTGCTGATGTCACGAGGTACCCGTAA 100
QY 365 AACCGACAGCATTTCTTGATGTGTGGCTCCGGGTGTGGTTGACCGTGCATATCA 424
Db 99 AACGCGAAGCCITTCGCCCGCGGCTGGACCTGTGCAATACCGTACATACCCA 40
QY 425 GCGCATTTGAAAAGACCGATCCCTGTTCCATCC 460
Db 39 GCGCATACGTCACGTCGATCCGATCCGTTATTATCC 4

RESULT 13
AA545747/c
LOCUS
DEFINITION
HBM5F1B4-REV Human Bone Marrow Stromal Fibroblast Homo sapiens cDNA
clone HBM5F1B4 5', mRNA sequence.
ACCESSION
AA545747
VERSION
AA545747.1 GI:2307026
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 354)
Jia, L., Robey, P., Young, M., Schuler, G., Powell, J., Yang, L.,
Lennon, G., Hillier, L., Allen, M., Bowles, L., Geisel, S., Kucaba, T.,
Marra, M., Martin, J., Steptoe, M., Tan, F., Theising, B., Bowers, Y.,
Wylie, T., Waterston, R., Wilson, R. and Francomano, C.
WashU-MGB/ANHRI EST Project
Unpublished (1997)
Contact: Libin Jia
Medical Genetics Branch
National Human Genome Research Institute
10/10C101, 9000 Rockville Pike, Bethesda, MD 20892-1267, USA
Tel: 301-402-4877

FEATURES
source

1..354
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HBM5F1B4"
/sex="Male and Female"
/tissue_type="bone marrow"
/cell_type="stromal fibroblast"
/dev_stage="mixed"
/lab_host="XL1-Blue"
/clone_lib="Human Bone Marrow Stromal Fibroblast"
/note="Vector: pBluescript; Site 1: EcoRI; Site 2: XhoI"

ORIGIN

Query Match 3.5%; Score 46.2; DB 9; Length 354;
Best Local Similarity 53.8%; Pred. No. 0.0057;
Matches 93; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
QY 696 AAGTGCACCTCAGTGGGCACTGCGTTATCATGCATTCGATGGTGAATCTTCTATTA 755
Db 174 ATGINTCATTAACCGTGGGTAGCCTCGCATCAATGCTGACGAGATATTTCCTG 115
QY 756 CAAAGCGCAAGCCATGCCAGAGTTGCTTGGCAACGGCTAAAGGGGGGAGAAATGG 815
Db 114 CAACAGCACAGGAATCGGAGCGGGTGGGAGGATCACCGATTACACCACTGG 55
QY 816 GTATCCTTATTGTCATTACATAACCGCAATTTAATTTAATGCGAAAACACC 868
Db 54 AACACCTTGCTAAGTTGCTAACCGCAATTTATTGCTACAACGACGCGC 2

RESULT 14
BX381961/c

LOCUS
DEFINITION
BX381961 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CSOD1072YF05 3-PRIME, mRNA sequence.
ACCESSION
BX381961
VERSION
BX381961.1 GI:30453007
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1201)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. Contact: Feng Liang Email: fliang@lifetech.com URL:
<http://fulllength.invitrogen.com/> Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID: CSOD1072CC03NP1.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. Contact: Feng Liang Email: fliang@lifetech.com URL:
<http://fulllength.invitrogen.com/> Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID: CSOD1072CC03NP1.

FEATURES
source

1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSOD1072YF05"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 3.3%; Score 44; DB 13; Length 1201;
Best Local Similarity 4.5%; Pred.No.0.057;
Matches 31; Conservative 245; Mismatches 418; Indels 1; Gaps 1;
QY 53 GCGGATTCGGCTACTTACTGCGCGGTAGCGCGAGCCATCGGGTATATTAGAAC 112
DB 1087 GKKGCCSKGKVKMGBMNKKQMSKVMGMMKMBKXMGKKKKKKKKKBBXTVK 1028
QY 113 GTGTGTTATTGTAGTCGCGATGTTGCTGCGCCGCAACAAACAGCGAGCTATG-A 171
DB 1027 KKTIVKKBMTGKTKKKKKKKMGMBMVKVMGKGVHKGKBMVBGKGMVMTGKM 968
QY 172 ATGATGTTACGCGAGATGAAGTGGCCACATGCGCGGTAAAGCGGGGTATTAAACGCCAC 231
DB 967 VKKBMGGKMKMKMBGMVKKTKMMVMBKMKMKMMVKKVMGKKKKMMKMMMM 908
QY 232 GTGTGCGGAGTGTGTCATGATGCGGGGGTATTATGCTGATTAATCTTCGCGACCTTG 291
DB 907 KKGKGMKGNMVKKQMKMKKGGTCKGKKKKKKKKKKKKKKKKKKKKKK 848
QY 292 GTTGTGAGCGCGGATGTCGCGAGAGGGGGGTATATGCAAGCAGATATCGATC 351
DB 847 MYKGMKGMKVKYKGMNKKKMGKGBKCMVKMMVMGKGMGKGMVMGMSK 788
QY 352 AACGTACCGGTTAACCGGACAGGCAATTTCTTGATGCTGCTCCGGGGTGTGTTGA 411
DB 787 MKMMNN 728
QY 412 CGTGCATAATCAGCGCGATTTGAAAGACCGATCCCTGTTCCATCGGTAGAGACTG 471
DB 727 NNANNMTNN 668
QY 472 GCGTGTGTAAGTTAGACACGACAAAGAGATTAAGAGATTAAGAGATTAAGAG 531
DB 667 NNN 608
QY 532 CGTTAGATACGCTGAGCGAGCGCTACGCTAAACCTTTTCCAGATGCGGAGCGTGA 591
DB 607 NNN 548
QY 592 ATTTTCGCGCTTCTCCTTATTGCAATCTTTGCAACAGCAAGGAAAGATGTTG 651
DB 547 KQNMKNMNTKQNMKNMNTKQNMKNMNTKQNMKNMNTKQNMKNMNTKQNMKN 488
QY 652 CCCATTTCGCGCAATGAGTTAAGTTAAGTTAAGAGAGGACAAAGTGCCTCAGT 711
DB 487 MNMNN 428
QY 712 GGCACCTGCGGTTATCATGACATTCGCTGAAATC 746
DB 427 KMMNNMKNMNN 393

RESULT 15
CNS00KK2/c
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence T7 end of BAC:
BACR17F22 of RPI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION
AL077673
VERSION
AL077673.1
KEYWORDS
GI:4957249
SOURCE
Drosophila melanogaster (fruit fly)
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

FEATURES
source
1..1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR17F22"
/clone_lib="RPI-98"
/notes="end : T7"

ORIGIN
Query Match 3.3%; Score 43.2; DB 29; Length 1101;
Best Local Similarity 10.1%; Pred.No.0.098;
Matches 36; Conservative 161; Mismatches 160; Indels 0; Gaps 0;

QY 82 GCGCGGAGCATCGGCTATATCTTTAGAACGCTGCTGTTTATTTAGTGGCATGCTGT 141
DB 1090 KCHNKK 1031
QY 142 CGTTCGCGGACCAACAGCAGCTTATGATGATGTTAGCCAGATAGTGCCACAAT 201
DB 1030 MKKKKKMMKTNBMMNKKDCNBTDMTKMCTMTTKTKKACMKMKMCTACCTBCTCKK 971
QY 202 GCGCGGTAAAGCGGGGTATTTAAACGCGCTGCTGCGAGTTGGTCACATTCATGGGG 261
DB 970 KMKKCTTKCKGCKGCKTKTKTKCHNMTKSTCBMTMTKVKGMATKMMMGCKGA 911
QY 262 GGTTTATGCTGATTAATCTTTTCGAGCCTTGGTTTGTAGCGGGGATGTCGCGAGAG 321
DB 910 CKDKKKMBKMDMKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 851
QY 322 GGGGGGTATGACAGGAGATATCGATCAACGCTACCGCTTAAACGAGACAGCATTC 381
DB 850 KKKKTKVKMMTKKMMNKKMMYKKKAMCKYCKDCMCKCKCKCKMMKMMKMMKMM 791
QY 382 TTGATGCTGCTGCTCCGGGGTGGTTTGAACCGTGATATCAGGGCGATTTGAAA 438
DB 790 TMTKVVMTK 734

Search completed: May 3, 2004, 03:54:06
Job time : 2502.04 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 30, 2004, 12:30:35 ; Search time 53.7805 Seconds
(without alignments)
2206.561 Million cell updates/sec

Title: US-10-021-723B-4

Perfect score: 2200

Sequence: 1 MSVLNVRVLSGLVLMISGL.....NPAGIISVAVACENNGDDK 420

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

1: Geneseq1980s.*

2: Geneseq1990s.*

3: Geneseq2000s.*

4: Geneseq2001s.*

5: Geneseq2002s.*

6: Geneseq2003as.*

7: Geneseq2003bs.*

8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2200	100.0	420	5 ABP51930	Abp51930 Yersinia
2	2193	99.7	441	5 ABP51934	Abp51934 Yersinia
3	2183	99.2	441	6 ABU49868	Abu49868 Protein e
4	2175	98.9	441	5 ABP51929	Abp51929 Yersinia
5	957.5	43.5	431	6 AAE37853	Aae37853 pNOV4054
6	957.5	43.5	437	6 AAE37854	Aae37854 pNOV4058
7	948.5	43.1	436	7 ADC87743	Adc87743 Modified
8	943	42.9	410	5 ABP51937	Abp51937 Phytase p
9	943	42.9	412	6 AAE37851	Aae37851 Maize-opt
10	939.5	42.7	412	7 ABR82310	AbR82310 Nov9X phy
11	939.5	42.7	432	3 AAB36257	Aab36257 Lama2/APP
12	939.5	42.7	432	3 AAB36261	Aab36261 R15/APP
13	939.5	42.7	432	3 AAB36262	Aab36262 SV40/APP
14	939.5	42.7	432	3 AAB36259	Aab36259 R15/APP
15	939.5	42.7	432	3 AAB36258	Aab36258 R15/APP
16	939.5	42.7	432	3 AAB36263	Aab36263 Lama2/APP
17	939.5	42.7	432	3 AAB36260	Aab36260 R15/APP
18	939.5	42.7	432	3 AAU77775	Aau77775 Phytase a
19	939.5	42.7	432	5 AAE15807	Aae15807 Escherich
20	939.5	42.7	432	5 ABP53625	Abp53625 Acid phos
21	939.5	42.7	432	6 ADAL19450	Adal19450 E. coli K
22	939.5	42.7	432	6 ABU15427	Abu15427 Protein e
23	939.5	42.7	432	7 ADC87745	Adc87745 Escherich
24	938.5	42.7	432	4 AAE02631	Aae02631 E. coli a
25	938.5	42.7	432	4 AAE02634	Aae02634 E. coli a

26	932.5	42.4	440	4 AAB37892	Aab37892 Escherich
27	932.5	42.4	440	5 AAE22836	Aae22836 Escherich
28	932.5	42.4	440	5 AAE15806	Aae15806 Escherich
29	932.5	42.4	440	6 ADAL19446	Adal19446 E. coli B
30	931.5	42.3	432	4 AAE02635	Aae02635 E. coli a
31	928.5	42.2	432	3 AAY94753	Aay94753 E. coli a
32	928.5	42.2	432	4 AAE02632	Aae02632 E. coli a
33	928.5	42.2	432	6 ABR42161	AbR42161 Escherich
34	928.5	42.2	476	5 ABP51933	Abp51933 Yersinia
35	921.5	41.9	430	5 AAE15808	Aae15808 Escherich
36	921.5	41.9	430	6 ADAL19451	Adal19451 E. coli K
37	920.5	41.8	432	6 ABR42162	AbR42162 Escherich
38	860	39.1	423	2 AAY01513	Aay01513 An Escher
39	567	25.8	318	5 ABP51936	Abp51936 Consensus
40	504	22.9	409	5 ABP51932	Abp51932 Rhizobium
41	482	21.9	421	5 ABP51931	Abp51931 Yersinia
42	469	21.3	261	4 ABG25698	Abg25698 Novel hum
43	291	13.2	123	4 AAU23269	Aau23269 Novel hum
44	282	12.8	393	5 ABP51935	Abp51935 Appa phyt
45	265	12.0	144	4 ABG24256	Abg24256 Novel hum

ALIGNMENTS

RESULT 1
ABP51930
ID ABP51930 standard; protein; 420 AA.
XX AC
XX ABP51930;
DT 08-OCT-2002 (first entry)
XX Yersinia pestis phytase protein sequence SEQ ID NO:4.
DE Yersinia pestis; phytase; enzyme; E.C.3.1.3.8; phytate; foodstuff; feed.
XX Yersinia pestis;
XX Yersinia pestis.
XX OS
XX Yersinia pestis.
PN WO200248332-A2.
XX
PD 20-JUN-2002.
XX
PF 12-DEC-2001; 2001WO-US048774.
XX
PR 12-DEC-2000; 2000US-0255090P.
XX
XX (DIVE-) DIVERSA CORP.
XX Short J, Mathur EJ, Richardson T, Robertson D, Barton N;
XX WPI: 2002-583504/62.
DR N-PSDB; ABQ73843.
XX
PT Novel recombinant phytase protein and polynucleotide for improving
PT nutritional value of phytate-containing foodstuff, in animal feed and
PT feed supplements and to degrade excess phytase from environment or
PT sample.
XX Claim 49; Fig 5D; 208pp; English.
XX
CC The present invention describes an isolated phytase protein (I). (I) can
CC be used for improving the nutritional value of a phytate-containing
CC foodstuff, by contacting the phytate-containing foodstuff with (I), where
CC the enzyme catalyses the liberation of inorganic phosphate from the
CC phytate-containing foodstuff, and so improving the nutritive value of the
CC contacted foodstuff. The liberation of inorganic phosphate occurs prior
CC to or after the ingestion of phytate-containing foodstuff by a recipient
CC organism. Nucleotide sequences (II) encoding (I) can be used for
CC producing an animal feed, by transforming a plant, plant portion or plant
CC cell with a nucleic acid expression vector, comprising (II), culturing
CC the plant, plant portion or plant cell under conditions in which the
CC phytase protein is expressed and converting the plant, plant portion or

CC plant cell into a composition suitable for animal feed. The animal is
CC preferably a monogastric animal or a ruminant. (I) and (II) are useful in
CC animal feed and feed supplements as well as in treatments to degrade or
CC remove excess phytate from the environment or a sample. (I) reduces
CC phytate levels in animal manure and so reduces the phosphate pollution of
CC the environment. The present sequence represents a phytase from the
CC present invention
XX
SQ Sequence 420 AA;

Query Match 100.0%; Score 2200; DB 5; Length 420;
Best Local Similarity 100.0%; Pred. No. 1.8e-210;
Matches 420; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSVLENRVRLSGLVLMISGLAAITAPVAAPSPGYTLRRVILSRHGVRSPTKOTQLMNDV 60
Db 1 MSVLENRVRLSGLVLMISGLAAITAPVAAPSPGYTLRRVILSRHGVRSPTKOTQLMNDV 60
QY 61 TPDKWPQWPVKAGVLTFRGAEVLTLMGFFYDIFRSGLLAAGCPAEGGVYAQADIDQRT 120
Db 61 TPDKWPQWPVKAGVLTFRGAEVLTLMGFFYDIFRSGLLAAGCPAEGGVYAQADIDQRT 120
QY 121 RLGTQAFLDGVAPEGGLTVHNAQDLKKTDLPLFHPVETGVCKLNAQTDKAIERLGGPLD 180
Db 121 RLGTQAFLDGVAPEGGLTVHNAQDLKKTDLPLFHPVETGVCKLNAQTDKAIERLGGPLD 180
QY 181 TVSORYAKPQAQMGDVILNFAASPYCKSLQOQGTCTDFAHFAANVNNKEGTVKTLSGPL 240
Db 181 TVSORYAKPQAQMGDVILNFAASPYCKSLQOQGTCTDFAHFAANVNNKEGTVKTLSGPL 240
QY 241 ALSSTLGEIFLLQNAQAMPVAVQRLKGAENWVSLSLHNAQFNLMAKTPYIARHKGTEL 300
Db 241 ALSSTLGEIFLLQNAQAMPVAVQRLKGAENWVSLSLHNAQFNLMAKTPYIARHKGTEL 300
QY 301 LQIDTALTQLDAQOGKLPISQNRVFLFGGHDNTNIAAGMLGADWQLPEQPDNTPPG 360
Db 301 LQIDTALTQLDAQOGKLPISQNRVFLFGGHDNTNIAAGMLGADWQLPEQPDNTPPG 360
QY 361 GGLVFELWQPNQPNHQRVAVVMFYQTMQDLRNEAKLDLKNPAGIISVAVACENNGDDK 420
Db 361 GGLVFELWQPNQPNHQRVAVVMFYQTMQDLRNEAKLDLKNPAGIISVAVACENNGDDK 420

RESULT 2
ABP51934
ID ABP51934 standard; protein; 441 AA.
XX AC ABP51934;
XX DT 08-OCT-2002 (first entry)
XX DE Yersinia pestis phytase protein sequence.
XX KW Yersinia pestis; phytase; enzyme; E.C.3.1.3.8; phytate; foodstuff; feed.
XX OS Yersinia pestis.
XX PN W0200248332-A2.
XX PD 20-JUN-2002.
XX PF 12-DEC-2001; 2001WO-US048774.
XX PR 12-DEC-2000; 2000US-0255090P.
XX PA (DIVE-) DIVERSA CORP.
XX PI Short J, Mathur EJ, Richardson T, Robertson D, Barton N;
XX WPI: 2002-583504/82.
XX DR N-PSDB; ABQ73847.
XX PT Novel recombinant phytase protein and polynucleotide for improving

PT nutritional value of phytate-containing foodstuff, in animal feed and
PT feed supplements and to degrade excess phytase from environment or
PT sample.
XX
XX Claim 57; Fig 5L; 208pp; English.

CC The present invention describes an isolated phytase protein (I). (I) can
CC be used for improving the nutritional value of a phytate-containing
CC foodstuff, by contacting the phytate-containing foodstuff with (I), where
CC the enzyme catalyses the liberation of inorganic phosphate from the
CC phytate-containing foodstuff, and so improving the nutritive value of the
CC contacted foodstuff. The liberation of inorganic phosphate occurs prior
CC to or after the ingestion of phytate-containing foodstuff by a recipient
CC organism. Nucleotide sequences (II) encoding (I) can be used for
CC producing an animal feed, by transforming a plant, plant portion or plant
CC cell with a nucleic acid, by transfection vector, comprising (II), culturing
CC the plant, plant portion or plant cell under conditions in which the
CC phytase protein is expressed and converting the plant, plant portion or
CC plant cell into a composition suitable for animal feed. The animal is
CC preferably a monogastric animal or a ruminant. (I) and (II) are useful in
CC animal feed and feed supplements as well as in treatments to degrade or
CC remove excess phytate from the environment or a sample. (I) reduces
CC phytate levels in animal manure and so reduces the phosphate pollution of
CC the environment. The present sequence represents a phytase from the
CC present invention
XX
SQ Sequence 441 AA;

Query Match 99.7%; Score 2193; DB 5; Length 441;
Best Local Similarity 99.8%; Pred. No. 9.9e-210;
Matches 419; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MSVLENRVRLSGLVLMISGLAAITAPVAAPSPGYTLRRVILSRHGVRSPTKOTQLMNDV 60
Db 1 MSVLENRVRLSGLVLMISGLAAITAPVAAPSPGYTLRRVILSRHGVRSPTKOTQLMNDV 60
QY 61 TPDKWPQWPVKAGVLTFRGAEVLTLMGFFYDIFRSGLLAAGCPAEGGVYAQADIDQRT 120
Db 61 TPDKWPQWPVKAGVLTFRGAEVLTLMGFFYDIFRSGLLAAGCPAEGGVYAQADIDQRT 120
QY 121 RLGTQAFLDGVAPEGGLTVHNAQDLKKTDLPLFHPVETGVCKLNAQTDKAIERLGGPLD 180
Db 121 RLGTQAFLDGVAPEGGLTVHNAQDLKKTDLPLFHPVETGVCKLNAQTDKAIERLGGPLD 180
QY 181 TVSORYAKPQAQMGDVILNFAASPYCKSLQOQGTCTDFAHFAANVNNKEGTVKTLSGPL 240
Db 181 TVSORYAKPQAQMGDVILNFAASPYCKSLQOQGTCTDFAHFAANVNNKEGTVKTLSGPL 240
QY 241 ALSSTLGEIFLLQNAQAMPVAVQRLKGAENWVSLSLHNAQFNLMAKTPYIARHKGTEL 300
Db 241 ALSSTLGEIFLLQNAQAMPVAVQRLKGAENWVSLSLHNAQFNLMAKTPYIARHKGTEL 300
QY 301 LQIDTALTQLDAQOGKLPISQNRVFLFGGHDNTNIAAGMLGADWQLPEQPDNTPPG 360
Db 301 LQIDTALTQLDAQOGKLPISQNRVFLFGGHDNTNIAAGMLGADWQLPEQPDNTPPG 360
QY 361 GGLVFELWQPNQPNHQRVAVVMFYQTMQDLRNEAKLDLKNPAGIISVAVACENNGDDK 420
Db 361 GGLVFELWQPNQPNHQRVAVVMFYQTMQDLRNEAKLDLKNPAGIISVAVACENNGDDK 420

RESULT 3
ABU49868
ID ABU49868 standard; protein; 441 AA.
XX AC ABU49868;
XX DT 19-JUN-2003 (first entry)
XX DE Protein encoded by Prokaryotic essential gene #35395.
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX

OS Yersinia pestis.
XX
XX PN WO200277183-A2.
XX
XX PD 03-OCT-2002.
XX
XX PF 21-MAR-2002; 2002WO-US009107.
XX
XX PF 21-MAR-2001; 2001US-00815242.
XX PR 06-SEP-2001; 2001US-00948993.
XX PR 25-OCT-2001; 2001US-0342923P.
XX PR 08-FEB-2002; 2002US-00072851.
XX PR 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX PA
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KI, Zyskind JW;
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
XX DR WPI; 2003-0299926/02.
XX DR N-PSDB; ACA53738.
XX
XX New antiseptic nucleic acids, useful for identifying proteins or screening
XX PT for homologous nucleic acids required for cellular proliferation to
XX PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 77792; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX CC the 6213 antiseptic sequences given in the specification where expression
XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:
XX CC (1) a vector comprising a promoter operably linked to the nucleic acid
XX CC encoding a polypeptide whose expression is inhibited by the antiseptic
XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX CC polypeptide or its fragment whose expression is inhibited by the
XX CC antiseptic nucleic acid; (4) an antibody capable of specifically binding
XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX CC proliferation or the activity of a gene in an operon required for
XX CC proliferation; (7) identifying a compound that influences the activity of
XX CC the gene product or that has an activity against a biological pathway
XX CC required for proliferation, or that inhibits cellular proliferation; (8)
XX CC identifying a gene required for cellular proliferation or the biological
XX CC pathway in which a proliferation-required gene or its gene product lies
XX CC or a gene on which the test compound that inhibits proliferation of an
XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX CC compound's activity; (11) a culture comprising strains in which the gene
XX CC product is overexpressed or underexpressed; (12) determining the extent
XX CC to which each of the strains is present in a culture or collection of
XX CC strains; or (13) identifying the target of a compound that inhibits the
XX CC proliferation of an organism. The antiseptic nucleic acids are useful for
XX CC identifying proteins or screening for homologous nucleic acids required
XX CC for cellular proliferation to isolate candidate molecules for rational
XX CC drug discovery programs, or for screening homologous nucleic acids
XX CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
XX CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
XX CC the target prokaryotic essential genes. Note: The sequence data for this
XX CC patent did not form part of the printed specification, but was obtained
XX CC in electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 441 AA;
XX
XX Query Match 99.2%; Score 2183; DB 6; Length 441;
XX Best Local Similarity 99.3%; Pred. No. 9.8e-209;
XX Matches 417; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
XX
XX 1 MSVLENRVRLSGVLMGLAAITAPVAAPSPGYTLERVVILSRGVSPKQTQLMNDV 60
XX Db 1 MSVLENRVRLSGVLMGLAAITAPVAAPSPGYTLERVVILSRGVSPKQTQLMNDV 60
XX
XX 61 TPDKFWQWPVKAQYLTFRGALVTLMGFGYGYFRSLGALLAGCPAEGGVAAQIDORT 120
XX Db 61 TPDKFWQWPVKAQYLTFRGALVTLMGFGYGYFRSLGALLAGCPAEGGVAAQIDORT 120
XX
XX The present invention describes an isolated phytase protein (I). (I) can

CC be used for improving the nutritional value of a phytate-containing
CC foodstuff, by contacting the phytate-containing foodstuff with (I), where
CC the enzyme catalyzes the liberation of inorganic phosphate from the
CC phytate-containing foodstuff, and so improving the nutritive value of the
CC contacted foodstuff. The liberation of inorganic phosphate occurs prior
CC to or after the ingestion of phytate-containing foodstuff by a recipient
CC organism. Nucleotide sequences (II) encoding (I) can be used for
CC producing an animal feed, by transforming a plant, plant portion or plant
CC cell with a nucleic acid expression vector, comprising (II), culturing
CC the plant, plant portion or plant cell under conditions in which the
CC phytase protein is expressed and converting the plant, plant portion or
CC plant cell into a composition suitable for animal feed. The animal is
CC preferably a monogastric animal or a ruminant. (I) and (II) are useful in
CC animal feed and feed supplements as well as in treatments to degrade or
CC remove excess phytate from the environment or a sample. (I) reduces
CC phytate levels in animal manure and so reduces the phosphate pollution of
CC the environment. The present sequence represents a phytase from the
CC present invention
XX
SQ Sequence 441 AA;

Query Match 98.9%; Score 2175; DB 5; Length 441;
Best Local Similarity 99.0%; Pred. No. 6.2e-208;
Matches 416; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 MSVLENRVLSGLVLMISGLAATAPVAAPSGYTLRVVLSRHGVRSPKOTQLMNDV 60
DB 1 MSVLENRVLSGLVLMISGLAATAPVAAPSGYTLRVVLSRHGVRSPKOTQLMNDV 60
QY 61 TPKWQWPVKAGYLTFRGAEVLTLMGGFYGYFRSLGLLAAGCPAEGGVYQAADIDQRT 120
DB 61 TPKWQWPVKAGYLTFRGAEVLTLMGGFYGYFRSLGLLAAGCPAEGGVYQAADIDQRT 120
QY 121 RLGTQAPLDGAVPGCGLTVHNOADLKKTDPLFHPVETGVCKLDNAOTDKAIEERLGGPLD 180
DB 121 RLGTQAPLDGAVPGCGLTVHNOADLKKTDPLFHPVETGVCKLDNAOTDKAIEERLGGPLD 180
QY 181 TVSQRYAKPPAQMAGDVLNPAASPYCKSLQQQKTCDFAHFAAENNVNKEGTVTLGGSL 240
DB 181 TVSQRYAKPPAQMAGDVLNPAASPYCKSLQQQKTCDFAHFAAENNVNKEGTVTLGGSL 240
QY 241 ALSTLGEIFLLQNAQAMPEVAVQRLKGAENWVLSLHNAQFNLMKATPYIARHKGTP 300
DB 241 ALSTLGEIFLLQNAQAMPEVAVQRLKGAENWVLSLHNAQFNLMKATPYIARHKGTP 300
QY 301 LQIDTALTQLDAQGGKLPISAQNRVFLGGHDTNIAIAGMLGADWQLEPQDNTPPG 360
DB 301 LQIDTALTQLDAQGGKLPISAQNRVFLGGHDTNIAIAGMLGADWQLEPQDNTPPG 360
QY 361 GGLVFEWQPNDRHQRVAVQMFYQTDQLRNAEKLDLKNPAGIISVAVAGCENGGDK 420
DB 361 GGLVFEWQPNDRHQRVAVQMFYQTDQLRNAEKLDLKNPAGIISVAVAGCENGGDK 420

RESULT 5
AAE37853
ID AAE37853 standard; protein; 431 AA.
XX
AC AAE37853;
XX
DT 06-NOV-2003 (first entry)
XX
DE pNOV4054 phytase fusion protein.
XX
KW Thermotolerant phytase; weight gain; animals fed diet; grain processing;
KW nutritive value; transformed plant; anabolic; chimeric; maize.
XX
OS Chimeric - Zea mays.
OS Chimeric - Unidentified.
XX
FH Key Location/Qualifiers
FT Peptide 1..19
FT /label= Signal_peptide

Protein 20. .431
/note= "Mature pNOV4054 phytase fusion protein"
WO2003057248-A1.
17-JUN-2003.
30-DEC-2002; 2002WO-US041787.
28-DEC-2001; 2001US-0344476P.
(SYGN) SYNGENTA PARTICIPATIONS AG.
Lanahan MB, Betts S;
WPI; 2003-607980/57.
N-PSDB; RAD57148.
Preparing a thermotolerant phytase for preparing animal feed or human
food by expressing in a plant cell an expression cassette comprising a
promoter operably linked to a nucleic acid molecule encoding a
thermotolerant phytase.
Claim 63; Page 90; 157pp; English.
The invention relates to a method for preparing a thermotolerant phytase.
The method comprises expressing in a plant cell an expression cassette
comprising a promoter operably linked to a nucleic acid molecule encoding
a thermotolerant phytase which retains at least 40% activity after 30
minutes at 60plusoc and has a specific activity of greater than 200 U/mg
at pH 4.5 and 37plusoc. The method is useful for preparing a
thermotolerant phytase for preparing animal feed or human food. The
invention is useful for reducing the feed conversion ratio and increasing
weight gain, improving reducing feed conversion ratios or increasing
weight gain of animals fed diets with inorganic phosphate at levels below
0.45%, minimising dietary requirements of phosphorus in an animal,
enhancing the utilisation of phosphorus present in animal feed, enhancing
organic phosphorus utilisation from organic phosphorus sources in animal
feed, decreasing the phosphate levels in excreta from an animal,
improving the processing of grain, improving the nutritive value of
processed grain product or a method of processing grain, improving the
nutritive value of animal feed and human food, and preparing a
transformed plant which expresses a thermotolerant phytase. The present
sequence is pNOV4054 phytase fusion protein used in the exemplification
of the invention
SQ Sequence 431 AA;

Query Match 43.5%; Score 957.5; DB 6; Length 431;
Best Local Similarity 48.3%; Pred. No. 2.4e-86;
Matches 199; Conservative 63; Mismatches 137; Indels 13; Gaps 7;
QY 7 RVLISGLVLMISGLAATAPVAAPSGYTLRVVLSRHGVRSPKOTQLMNDVTPDKWP 66
DB 2 RVLVVALAL-LALAAATSAQAQSEPE-LKLESVVIVSRHGVRAPTKATQLMODVTPDAMP 59
QY 67 QWPVKAGYLTFRGAEVLTLMGGFYGYFRSLGILL-AAGCPAEGGVYQAADIDQRTLTGQ 125
DB 60 TWPVKLGELTPRGELIATYLGHWYRQRLVADGLLPKGCPSQGVIAIDVDETRKTGE 119
QY 126 AFLDGVAPCGGLTVHNOADLKKTDPLFHPVETGVCKLDNAOTDKAIEERLGGPLDTSVQR 185
DB 120 AFAAGLAPDCAITVHTQADTSSPDPLENPLKTVGVCOLDNANVTDAILERAGGSIAFTGH 179
QY 186 YAKPPAQMAGDVLNPAASPYCKSLQQQKTCDFAHFAAENNVNKEGTVTLGGSLPALSST 245
DB 180 YQAFRELRVFLNFPQSNLCKREKQDECSLTQALPSCLKVSAD--CVSITGAVSLASM 237
QY 246 LGEIFLLQNAQAMPEVAVQRLKGAENWVLSLHNAQFNLMKATPYIARHKGTPLOQID 305
DB 238 LTEIFLLQQAQMGPEPGWGRITDSHOWNTLLSLHNAQFDLLQRTPEVARSATPLLDLIK 297
QY 306 TALT---LQIDAQGGKLPISAQNRVFLGGHDTNIAIAGMLGADWQLEPQDNTPPGG 362

Db 298 TALTTHPPKQKQYGVTLPTS-----VLFTAGHDTNLNAGGALNLTWLPQPDNTPPGGE 353

QY 363 LVFELWQNPDRHQRYVAVKMFYQTMQDLRNLAEKLDLKNPAGIISVAVAGCE 414

Db 354 LVFERWRRLSDNSQWIQVSLVFQTLQQRDXTPLSL-NTPPGVKLTLAGCE 404

RESULT 6

AAE37854

ID AAE37854 standard; protein; 437 AA.

AC AAE37854;

XX

DT 06-NOV-2003 (first entry)

XX

DE pNOV4058 phytase fusion protein.

XX

XX Thermotolerant phytase; weight gain; animals fed diet; grain processing;

KW nutritive value; transformed plant; anabolic; chimeric; maize.

XX

XX Chimeric - Zea mays.

OS Chimeric - Unidentified.

XX

XX Key Location/Qualifiers

FT Peptide 1..19

FT /label= Signal_peptide

FT Protein 20..437

FT /note= "Mature pNOV4058 phytase fusion protein"

FT Region 432..437

FT /note= "SEXDEL signal"

XX

XX WO2003057248-A1.

PN

XX

PD 17-JUL-2003.

XX

PF 30-DEC-2002; 2002WO-US041787.

XX

PR 28-DEC-2001; 2001US-0344476P.

XX

XX (SYGN) SYNGENTA PARTICIPATIONS AG.

PA

XX Lanahan MB, Betts S;

PI

XX WPI: 2003-607980/57.

DR

DR N-PSDB; AAD57149.

XX

XX Preparing a thermotolerant phytase for preparing animal feed or human

PT food by expressing in a plant cell an expression cassette comprising a

PT promoter operably linked to a nucleic acid molecule encoding a

PT thermotolerant phytase.

XX

PS Example 1; Page 91; 157pp; English.

XX

CC The invention relates to a method for preparing a thermotolerant phytase.

CC The method comprises expressing in a plant cell an expression cassette

CC comprising a promoter operably linked to a nucleic acid molecule encoding

CC a thermotolerant phytase which retains at least 40% activity after 30

CC minutes at 60pluoc and has a specific activity of greater than 200 U/mg

CC at pH 4.5 and 37pluoc. The method is useful for preparing a

CC thermotolerant phytase for preparing animal feed or human food. The

CC invention is useful for reducing the feed conversion ratio and increasing

CC weight gain, improving reduced feed conversion ratios or increasing

CC weight gain of animals fed diets with inorganic phosphate at levels below

CC 0.45%, minimising dietary requirements of phosphorus in an animal,

CC enhancing the utilisation of phosphorus present in animal feed, enhancing

CC organic phosphorus utilisation from organic phosphorus sources in animal

CC feed, decreasing the phosphate levels in excreta from an animal,

CC improving the processing of grain, improving the nutritive value of

CC processed grain product or a method of processing grain, improving the

CC nutritive value of animal feed and human food, and preparing a

CC transformed plant which expresses a thermotolerant phytase. The present

CC transference is pNOV4058 phytase fusion protein used in the exemplification

CC of the invention

XX

SQ Sequence 437 AA;

Query Match 43.5%; Score 957.5; DB 6; Length 437;

Best Local Similarity 48.3%; Pred. No. 2.5e-86;

Matches 199; Conservative 63; Mismatches 137; Indels 13; Gaps 7;

QY 7 RVLSGLVLMGLAAITAPVAAPSGYTLERVILSRHGVRSPKQTKQTMNDVTPDKWP 66

Db 2 RVLLVALAL-LALAASATSAQSEPE-LKLESVIVSRHGVRAPTKATQTMQDVTDPDWP 59

QY 67 QNPVKAGYLTTPRGAEALVTLMGGFYGDYFRSLGLL-AAGCPAEGGVYAQADIDQRTLTGQ 125

Db 60 TWPVKLGELTPRGGLIAYLGHVWRQRLVADGLLPGCGPQSGQVAIIADVDEKTEKGE 119

QY 126 AFLDGVAPCGGLTVRNOADLKTDDPLFHPVETGVCKLDNAQTDKALSERLGGPLDVSOR 185

Db 120 AFAAGLAPDCAITVHTQADTSSPDPLFNLKTKGVCOQDNLNVTDAILEPAGGSIADFTGH 179

QY 186 YAKPPAQMGGVNLNFAASPYCKSLQOQKCTDPAHFAANEVNVNKEGTVTLSPGLALSST 245

Db 180 YQTAFRELERLVNFQSNLCLKREKQDESCSLTQALPSELKVSAD--CVSLTAGVSLASM 237

QY 246 LGEIFLLQNAQAMPVAVQRLKGAENWVLSLLHNAQFNLMAKTPYIAHKGTPLLQQID 305

Db 238 LTBIFLLQQAQGMPEPGWGRITDSDHWNLTLLSHNAQFDLLQRTPEVARSRATPLLDLIK 297

QY 306 TALT---LQDQAQOKLPISAQNRVFLGCHDTNIANIAGMLGADWQLEQPDNTPPGGG 362

Db 298 TALTTHPPKQKQYGVTLPTS-----VLFTAGHDTNLNAGGALNLTWLPQPDNTPPGGE 353

QY 363 LVFELWQNPDRHQRYVAVKMFYQTMQDLRNLAEKLDLKNPAGIISVAVAGCE 414

Db 354 LVFERWRRLSDNSQWIQVSLVFQTLQQRDXTPLSL-NTPPGVKLTLAGCE 404

RESULT 7

ADC87743

ID ADC87743 standard; protein; 436 AA.

XX

AC ADC87743;

XX

DT 01-JAN-2004 (first entry)

XX

DE Modified Escherichia coli phytase.

XX

KW Phytase; food supplement; enzyme delivery matrix; soybean meal;

KW thermotolerance; thermostability; kernel; phytate;

KW myo-inositol-hexaphosphate; inositol; inorganic phosphate;

KW thermotolerant; feed value; digestion; enzyme.

XX

OS Synthetic.

OS Escherichia coli.

XX

XX US2003103958-A1.

XX

PD 05-JUN-2003.

XX

PF 24-MAY-2002; 2002US-00156660.

XX

XX 13-AUG-1997; 97US-00910798.

PR 01-MAR-1999; 99US-00259214.

PR 13-APR-1999; 99US-00291931.

PR 25-MAY-1999; 99US-00318528.

PR 25-MAY-2000; 2000US-00580515.

PR 24-MAY-2001; 2001US-00866379.

XX

XX (DIVE-) DIVERSA CORP.

PA

XX Short JM, Kretz K, Gray KA, Barton NR, Garrett JB, O'donoghue E;

PI Mathur EU;

XX

DR WPI; 2003-787039/74.
DR N-PSDB; ADC87742.
PT New nucleic acid encoding a polypeptide having phytase activity, useful
PT in improving the feeding value of phytate rich ingredients or as an aid
PT in phytate digestion.
XX
XX
PS Claim 60; SEQ ID NO 2; 113pp; English.
XX
XX The invention discloses a new isolated or recombinant nucleic acid which
CC encodes a polypeptide having a phytase activity. Also claimed is a
CC nucleic acid probe, an amplification primer sequence pair, an expression
CC cassette comprising the nucleic acid, a vector comprising the nucleic
CC acid, a transgenic non-human animal or plant, or its seed, comprising the
CC nucleic acid, an antisense oligonucleotide, inhibiting the translation of
CC a phytase message in a cell, a heterodimer comprising the polypeptide and
CC a second domain, an array comprising immobilised polypeptide or nucleic
CC acid, a hybridoma comprising an antibody that specifically binds to the
CC polypeptide, a food supplement for an animal, an edible enzyme delivery
CC matrix, an edible pellet comprising a granule edible carrier and the
CC polypeptide, a feed composition, a soybean meal, isolating or identifying
CC the polypeptide, making an anti-phytase antibody, producing a recombinant
CC polypeptide, determining whether a compound binds to the polypeptide,
CC identifying a modulator, whole cell engineering of new or modified
CC phenotypes by using real-time metabolic flux analysis, increasing
CC thermotolerance or thermostability of the phytase polypeptide, increasing
CC the resistance of the phytase polypeptide to enzymatic inactivation in a
CC digestive system and processing of corn and sorghum kernels. The phytase
CC activity comprises catalysis of phytate (myo-inositol-hexaphosphate) to
CC inositol and inorganic phosphate and is thermotolerant. The nucleic acid
CC is useful in improving the feeding value of phytate rich ingredients or
CC as an aid in phytate digestion. The sequence presented is the modified
CC Escherichia coli phytase.
XX
SQ Sequence 436 AA;
Query Match 43.1%; Score 948.5; DB 7; Length 436;
Best Local Similarity 48.2%; Pred. No. 1.9e-85;
Matches 197; Conservative 62; Mismatches 135; Indels 15; Gaps 7;
QY 13 LVMLSLGIAITAPVA---REPSGYTLERVVILSRHGVRPTQTOLMNDVTPDKPQWP 69
DB 5 LIPFLSLIPUTPOSAFAQSEPE-LKLESVVIVSRHGVRAPTATQLMQDVTDAWPTWP 63
QY 70 VKAGYLPRAELVTLMGFGYDFRSIGLL-AAGCPAEGGVYAQADIDORTLTQOAF 128
DB 64 VKLGELTPRGELIAYLGHYWRQLVADGLLPKCGCPQSGQVAILADVDERTRKTGEAFA 123
QY 129 DGVPAGGCLTVHNOADLKKTDPLPHVETGVCKLDNAQTOKAIEERLGGPLDVTVSQRYAK 188
DB 124 AGLAPDCAITVHTQADTSSPDPLFNPLKTGVCQLDANVTDAILERAGGSIADFTGHYQT 183
QY 189 PFAQMGVNLNFAAPYCKSLQQQKTCDFAHFAANEVNVNKEGKVTLSGLPLALSSTLGE 248
DB 184 AFRLERLVNPQSNCLKREKQDESCLTQALPSELKVSAD--CVSLTGAVSLASMLTE 241
QY 249 IFLLQNAQAMPEVAWQRLKGAENVVLSLHNAQFNLMKPTIARHKGTPFLQQIDTAL 308
DB 242 IFLLQQAQMPPEPGWGRITDQHWNTLLSHNAQFDLLQRTPEVARSRAEFLDLKNTAL 301
QY 309 T---LQIDAQCKLPISAQNEVLFGGHDNTNIAIAGMLGADWOLPEQDNTPPGGGLVF 365
DB 302 TPHPQQAQVGTIPTS-----VLFIAHDNTNLMGALSELNWTLPQDNTPPGGGLVF 357
QY 366 ELWQNPNDHORYAVKMYQTMQDLRNAEKLDDLKQNPAGIISVAVAGE 414
DB 358 ERWRRLSDNSQWISQVSLVFQTLQMRDKTPLSL-NTPPGEVKLTLAGCE 405
RESULT 8
ABP51937
ID ABP51937 standard; protein; 410 AA.
XX

AC ABP51937;
XX
XX 08-OCT-2002 (first entry)
XX
XX Phytase protein sequence.
DE
XX
XX Phytase; enzyme; E.C.3.1.3.8; phytate; foodstuff; feed.
KW
XX
XX Unidentified.
OS
XX
XX WO200248332-A2.
PN
XX
XX 20-JUN-2002.
XX
XX 12-DEC-2001; 2001WO-US048774.
PF
XX
XX 12-DEC-2000; 2000US-0255090P.
PR
XX
XX (DIVE-) DIVERSA CORP.
PA
XX
XX Short J, Mathur EJ, Richardson T, Robertson D, Barton N;
PI
XX
XX WPI; 2002-583504/62.
DR
XX
XX Novel recombinant phytase protein and polynucleotide for improving
PT nutritional value of phytate-containing foodstuff, in animal feed and
PT feed supplements and to degrade excess phytase from environment or
PT sample.
XX
XX
PS Disclosure; Fig 9; 208pp; English.
XX
XX The present invention describes an isolated phytase protein (I). (I) can
CC be used for improving the nutritional value of a phytate-containing
CC foodstuff, by contacting the phytate-containing foodstuff with (I), where
CC the enzyme catalyses the liberation of inorganic phosphate from the
CC phytate-containing foodstuff, and so improving the nutritive value of the
CC contacted foodstuff. The liberation of inorganic phosphate occurs prior
CC to or after the ingestion of phytate-containing foodstuff by a recipient
CC organism. Nucleotide sequences (II) encoding (I) can be used for
CC producing an animal feed, by transforming a plant, plant portion or plant
CC cell with a nucleic acid, expression vector, comprising (II), culturing
CC the plant, plant portion or plant cell under conditions in which the
CC phytase protein is expressed and converting the plant, plant portion or
CC plant cell into a composition suitable for animal feed. The animal is
CC preferably a monogastric animal or a ruminant. (I) and (II) are useful in
CC animal feed and feed supplements as well as in treatments to degrade or
CC remove excess phytate from the environment or a sample. (I) reduces
CC phytate levels in animal manure and so reduces the phosphate pollution of
CC the environment. The present sequence represents a phytase from the
CC present invention
XX
SQ Sequence 410 AA;
Query Match 42.9%; Score 943; DB 5; Length 410;
Best Local Similarity 49.0%; Pred. No. 6.2e-85;
Matches 191; Conservative 60; Mismatches 127; Indels 12; Gaps 6;
QY 29 ABPSGYTLERVVILSRHGVRPTQTOLMNDVTPDKPQWPKAGYLPRAELVTLMGG 88
DB 2 SEPE-LKLESVVIVSRHGVRAPTATQLMQDVTDAWPTWFKLGELTPRGELIAYLGH 60
QY 89 FYGDYFRSISGLL-AAGCPAEGGVYAQADIDORTLTQOAFLDGVAPGCLTVHNOADLKK 147
DB 61 YWRQLVADGLLPKCGCPQSGQVAILADVDERTRKTGEAFAAGLAPDCAITVHTQADTSS 120
QY 148 TDPLPHVETGVCKLDNAQTOKAIEERLGGPLDVTVSQRYAKPFAQMGVNLNFAAPYCKS 207
DB 121 PDPLFNPLKTGVCQLDANVTDAILERAGGSIADFTGHYQTAFRELERVLNFPQSNCLK 180
QY 208 LQQQKTCDFAHFAANEVNVNKEGKVTLSGLPLALSSTLGEIFLLQNAQAMPEVAWQRLK 267
DB 181 REKQDESCLTQALPSELKVSAD--CVSLTGAVSLASMLTEIFLLQQAQMPPEPGWGRIT 238

Db 64 VKLGWLTFRGGELIAYLGHYQRLVADGLLAKKGCPSQGVVAIIADVDERTKTEGAF 123
Qy 129 DGVAPGGLTVHNOADLKKTDPLFHPVETGVCKLDNAQTDKATEEBELGGPLDTVSQYAK 188
Db 124 AGLAPDCAITVHTQADTSSPDPLFNPLKTVGCOLDNANVTDAILSRAGGSIAADFTGHRQT 183
Qy 189 PFAQMGDVNLFAASPYCKSLQOQKTCDFAHFAANEVNVNKEGTVKTLSPGLALSSTLGE 248
Db 184 AFRELRLVNFQSNLCKREKQDESCSLTQALPSLKVSA--NVSLTGAVSLASMLTE 241
Qy 249 IFLIQNAQAMPEVAVQRLKGAENWVLSLSLHNAQFNLMAKTPYIAHKGTPPLQQIDTAL 308
Db 242 IFLQQAGMPEPGWGRITDSHOWNTLSLHNAQFYLQRTPEVARSRAATPLDLIKTAL 301
Qy 309 T---LQIDAQOQKLPISAQNRVFLGGHDTNINIANIAGMLGADWQLPEQPDNTPPGGGLVF 365
Db 302 TPHPQKQAGVTLPTS---VLFIAGHDTNLANLGALNLWTLPGQPDNTPPGGGLVF 357
Qy 366 ELWQNPNDHORYAVKMFYQTMQLRNAEKLDKNNPAGIISVAVAGE 414
Db 358 ERWRRLSDNSQWIOVSLVFQTLQOMRDKTPLSL-NTPPGEVKLTLAGCE 405
RESULT 14
AAB36259 standard; protein; 432 AA.
AC AAB36259;
XX
XX
DT 12-SEP-2003 (revised)
DT 20-FEB-2001 (first entry)
XX
DE R15/APPA plasmid translated sequence.
XX
KW Transgenic animal; salivary protein; phytase; phosphorus; animal growth;
KW environmental pollution; pig.
OS Rattus sp.
OS Escherichia coli.
OS Chimeric.
FN WO200064247-A1.
XX
XX
XX 02-NOV-2000.
XX
XX 20-APR-2000; 2000WO-CA000430.
XX
XX 23-APR-1999; 99US-0130508P.
XX
XX (UYGU-) UNIV GUELPH.
XX
XX Forsberg CW, Golovan S, Phillips JP;
XX
XX WPI; 2000-687245/67.
XX N-PSDB; AAC68296.
XX
PT Transgenic non-human animal for gastrointestinal tract specific
PT expression of a protein, preferably phytase, comprises a nucleic acid
PT sequence including a heterologous transgene construct encoding the
PT protein.
XX
XX
XX Disclosure; Fig 19; 152pp; English.
XX
CC The present invention provides transgenic animals which produce desired
CC proteins, in this case pigs which expresses phytase in the salivary
CC gland. Low phytase production levels result in phytate in the diet being
CC excreted and causing phosphorus contamination in water, as well as
CC reducing the growth of animals. The invention provides a number of
CC transgenes containing the E. coli APPA phytase coding sequence. (Updated
CC on 12-SEP-2003 to standardise OS field)
XX
SQ Sequence 432 AA;

Query Match 42.7%; Score 939.5; DB 3; Length 432;
Best Local Similarity 47.9%; Pred. No. 1.5e-84;
Matches 196; Conservative 61; Mismatches 137; Indels 15; Gaps 7;
Qy 13 LVLMISGLAIAITAPVA--AEPSYTLRVLRSRHSRSPKOTQOLMNDVTPDKWQWP 69
Db 5 LIPFLSLIPLTPOSATFAQSEPE-LKLESVVIVSRHGVRAPTKATQLMQDVTPDAMWTP 63
Qy 70 VKAGVLTFRGAEVLTLMGGFYDYFRSILGLIA-AGCPAEGGVAAQADIDORTLITGOAFL 128
Db 64 VKLGWLTFRGGELIAYLGHYQRLVADGLLAKKGCPSQGVVAIIADVDERTKTEGAF 123
Qy 129 DGVAPGGLTVHNOADLKKTDPLFHPVETGVCKLDNAQTDKATEEBELGGPLDTVSQYAK 188
Db 124 AGLAPDCAITVHTQADTSSPDPLFNPLKTVGCOLDNANVTDAILSRAGGSIAADFTGHRQT 183
Qy 189 PFAQMGDVNLFAASPYCKSLQOQKTCDFAHFAANEVNVNKEGTVKTLSPGLALSSTLGE 248
Db 184 AFRELRLVNFQSNLCKREKQDESCSLTQALPSLKVSA--NVSLTGAVSLASMLTE 241
Qy 249 IFLIQNAQAMPEVAVQRLKGAENWVLSLSLHNAQFNLMAKTPYIAHKGTPPLQQIDTAL 308
Db 242 IFLQQAGMPEPGWGRITDSHOWNTLSLHNAQFYLQRTPEVARSRAATPLDLIKTAL 301
Qy 309 T---LQIDAQOQKLPISAQNRVFLGGHDTNINIANIAGMLGADWQLPEQPDNTPPGGGLVF 365
Db 302 TPHPQKQAGVTLPTS---VLFIAGHDTNLANLGALNLWTLPGQPDNTPPGGGLVF 357
Qy 366 ELWQNPNDHORYAVKMFYQTMQLRNAEKLDKNNPAGIISVAVAGE 414
Db 358 ERWRRLSDNSQWIOVSLVFQTLQOMRDKTPLSL-NTPPGEVKLTLAGCE 405
RESULT 15
AAB36259
ID AAB36258 standard; protein; 432 AA.
XX
XX AAB36258;
XX
XX 12-SEP-2003 (revised)
XX 20-FEB-2001 (first entry)
XX
XX R15/APPA plasmid translated sequence.
XX
KW Transgenic animal; salivary protein; phytase; phosphorus; animal growth;
KW environmental pollution; pig.
XX
OS Rattus sp.
OS Escherichia coli.
OS Chimeric.
XX
XX WO200064247-A1.
XX
XX 02-NOV-2000.
XX
XX 20-APR-2000; 2000WO-CA000430.
XX
XX 23-APR-1999; 99US-0130508P.
XX
XX (UYGU-) UNIV GUELPH.
XX
XX Forsberg CW, Golovan S, Phillips JP;
XX
XX WPI; 2000-687245/67.
XX N-PSDB; AAC68296.
XX
PT Transgenic non-human animal for gastrointestinal tract specific
PT expression of a protein, preferably phytase, comprises a nucleic acid
PT sequence including a heterologous transgene construct encoding the
PT protein.
XX
XX
XX Disclosure; Fig 18; 152pp; English.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 30, 2004, 12:36:16 ; Search time 16.8293 Seconds
(without alignments)
1288.405 Million cell updates/sec

Title: US-10-021-723B-4

Perfect score: 2200
Sequence: 1 MSVLENRVRLSGLVLMISGL.....NPAGIISVAVAGCENNGDDK 420

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/prodata/2/iaa/5A_COMB.pep:*
- 2: /cgn2_6/prodata/2/iaa/5B_COMB.pep:*
- 3: /cgn2_6/prodata/2/iaa/6A_COMB.pep:*
- 4: /cgn2_6/prodata/2/iaa/6B_COMB.pep:*
- 5: /cgn2_6/prodata/2/iaa/PCUS_COMB.pep:*
- 6: /cgn2_6/prodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	932.5	42.4	440	3	US-09-259-214-2
2	932.5	42.4	440	3	US-09-318-528-2
3	932.5	42.4	440	3	US-09-291-931-2
4	928.5	42.2	433	4	US-09-540-149A-1
5	860	39.1	423	2	US-08-910-798-2
6	498	22.6	421	4	US-09-489-039A-7512
7	440.5	20.0	522	4	US-09-489-039A-13501
8	115	5.2	443	4	US-09-684-855-112
9	115	5.2	443	4	US-09-684-855-134
10	115	5.2	443	4	US-09-684-855-157
11	115	5.2	450	4	US-09-044-718-9
12	115	5.2	466	3	US-08-868-435-31
13	115	5.2	466	4	US-08-744-231-31
14	115	5.2	466	4	US-09-273-871A-13
15	99	4.5	2231	4	US-09-252-991A-21854
16	98.5	4.5	1257	4	US-09-336-987A-5
17	97.5	4.4	4928	3	US-09-032-991A-17290
18	97.5	4.4	4928	3	US-09-370-700-5
19	97.5	4.4	4928	4	US-09-603-207-5
20	96.5	4.4	947	4	US-09-543-681A-6556
21	94	4.3	449	3	US-08-680-506-7
22	94	4.3	660	4	US-09-107-532A-6715
23	93.5	4.2	504	4	US-09-543-681A-6521
24	93.5	4.2	1584	4	US-09-457-040B-27
25	92.5	4.2	439	4	US-09-543-681A-7293
26	91.5	4.2	394	6	5223418-2
27	91.5	4.2	529	1	US-08-548-509-2

28	91	4.1	452	4	US-09-489-039A-7220	Sequence 7220, Ap
29	91	4.1	506	1	US-08-313-075A-38	Sequence 38, Appl
30	91	4.1	506	3	US-08-606-505B-65	Sequence 65, Appl
31	91	4.1	506	3	US-09-616-590-65	Sequence 65, Appl
32	90.5	4.1	578	4	US-09-489-039A-7288	Sequence 7288, Ap
33	90	4.1	761	4	US-09-328-352-5650	Sequence 5650, Ap
34	90	4.1	1037	4	US-09-543-681A-7677	Sequence 7677, Ap
35	89.5	4.1	1056	4	US-09-595-684B-29	Sequence 29, Appl
36	89.5	4.1	1057	3	US-09-541-782-10	Sequence 10, Appl
37	89.5	4.1	1057	4	US-09-723-820-10	Sequence 10, Appl
38	89.5	4.1	1057	4	US-10-270-085-10	Sequence 10, Appl
39	89.5	4.1	1102	4	US-09-252-991A-31997	Sequence 31997, A
40	89.5	4.1	2284	4	US-09-252-991A-23547	Sequence 23547, A
41	89.5	4.1	3567	2	US-07-642-734C-4	Sequence 4, Appl
42	89.5	4.1	3567	3	US-08-439-009A-4	Sequence 4, Appl
43	89	4.0	508	4	US-09-126-420A-19	Sequence 19, Appl
44	88.5	4.0	613	4	US-09-252-991A-19955	Sequence 19955, A
45	88.5	4.0	1498	4	US-09-252-991A-31234	Sequence 31234, A

ALIGNMENTS

RESULT 1
US-09-259-214-2
; Sequence 2, Application US/09259214A
; Patent No. 6110719
; GENERAL INFORMATION:
; APPLICANT: Kretz, Keith
; TITLE OF INVENTION: NOVEL PHYTASE
; FILE REFERENCE: DIVER1370-1
; CURRENT APPLICATION NUMBER: US/09/259,214A
; CURRENT FILING DATE: 1999-03-01
; EARLIER APPLICATION NUMBER: 09/910,798
; EARLIER FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-259-214-2

Query Match	42.4%	Score	932.5	DB 3	Length	440			
Best Local Similarity	47.7%	Pred. No.	4.9e-92						
Matches	195	Conservative	61	Mismatches	138	Indels	15	Gaps	7
QY	13	LVLMLSLAIAITAPVA---AEPGTYLERVVILSRHGVRSPFTQTQLMNDVTPDKWPWP	69						
DB	5	LIPFLSILLPLTPQSAFAQSEPE-LKLESVIVSRHGVRAPTKATQLMQDVPDAMPWP	63						
QY	70	VKAGYLTPRCAELVTLMGFGVDYFRSLGLLA-AGCPAEGGVAAQADIDORTLTTGQAF	128						
DB	64	VKLGLWTPRGELIAYIGHYQRLVADGILLAKKGCPSQGVAAIADVDERTKTKGAPA	123						
QY	129	DGVAPGCGLTIVHNOADLKTKDPLFHPVETGVCKLDNAQTDKAIEERLGGPLDVTVSQRYAK	188						
DB	124	AGLAPDCAITVHTQADTSSPDPLFNPLKTVGCQLDANAVTDAILSRAGGSIAIDFTGHQT	183						
QY	189	PFAQMGDVLPFAAPSPCKSLQOQKCTCFAHFAANVNVNKEGTVKVTLSGPLALSSTIGE	248						
DB	184	AFRELERLVNFPQSNLCIKREKODECSLTQALPSELKVSAD--NVSLTGAVSLASMLTE	241						
QY	249	IFLLQNAQAPPEVAVQWRLKGAENWVSLLSLHNAQFNLMAKTPYIARHKGTPLLQOQIDTAL	308						
DB	242	IFLLQNAQAPPEVAVQWRLKGAENWVSLLSLHNAQFYLLOSTPEVARSRAATPLDLIMAAL	301						
QY	309	T---LQIDAQOQKLPISAQNRVLPLGHDNTNIAIAGMLGADWOLPEQONTTPGGGLVF	365						
DB	302	TPHPQKQAYGVTLPTS-----VLPIAGHDTNLNLGALBNMTLPQOPNTPPGGELVF	357						
QY	366	ELWQNPNDHORYVAVKMFYQTMQDLRNAEKLDLKNPNPAGIISVAVAGCE	414						

Db 358 ERWRRLSDNSQWIVQSLVFTLQMRDKTPLSL-NTPPGEVKLTLAGCE 405

RESULT 2

US-09-318-528-2
; Sequence 2, Application US/09318528
; Patent No. 6183740
; GENERAL INFORMATION:
; APPLICANT: Kretz, Keith
; TITLE OF INVENTION: NOVEL PHYTASE
; FILE REFERENCE: 09010/029003
; CURRENT APPLICATION NUMBER: US/09/318,528
; CURRENT FILING DATE: 1999-05-25
; EARLIER APPLICATION NUMBER: 09/291,931
; EARLIER FILING DATE: 1999-04-13
; EARLIER APPLICATION NUMBER: 08/910,798
; EARLIER FILING DATE: 1997-08-13
; EARLIER APPLICATION NUMBER: 09/259,214
; EARLIER FILING DATE: 1999-03-01
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-318-528-2

Query Match 42.4%; Score 932.5; DB 3; Length 440;
Best Local Similarity 47.7%; Pred. No. 4.9e-92;
Matches 195; Conservative 61; Mismatches 138; Indels 15; Gaps 7;
QY 13 LVLMLSLAAITAPVA---AEPSCYTLERVVILSRHGVRSPTKOTQLMNDVTPDKWPMP 69
Db 5 LIPFLSLIPLTPQSAFAQSEPE-LKLESVVIVSRHGVRAPTKATQLMQDVTDPDWTWP 63
QY 70 VKAGYLPFGAELVTLMGGFYGYFRSLGLLA-AGCPAEGGVYAQADIDORTLTGQAF 128
Db 64 VLGLWLTFRGGLTAYLGHYQORLVADGLLAKKGCPSQGVIAIDVDETRTKTGEAFA 123
QY 129 DGVAPEGGLTVHQAADLKTDPLFHPVETGVCKLDNAQTDKAIEERLGGPLDVTVSRYAK 188
Db 124 AGLAPDCAITVHTQADTSSPDPLFNPKTGVCQLDNANVTDAILSRAGGSIAFTGHROT 183
QY 189 PFAQMGDVLPFAASPYCKSLQOQKTCDFAHFAANEVNVNKEGTVKLTSGPLALSSTLGE 248
Db 184 AFRELRLVLPQSNLCKREKQDECSLTQALPSELKVSAD--NVSLTGAVSLASMLTE 241
QY 249 IFLQNAQAMPEVAWQRLKGAENVVSLSLHNAQFNLMAKTPYIARHKGTPLLQQIDTAL 308
Db 242 IFLQQAQGMPEPGWGRITDQSHQNTLLSLHNAQFYLLQRTPEVARSRATPDLDLMAAL 301
QY 309 T---LQDAQOQKLPISAQNRVFLGGHDTNIAAGMLGADWQLPEQPDNTPPGGGLVF 365
Db 302 TPHPPOQAQYGVTLPTS---VLFIAGHDTNLANLGGALELNTLPGQPDNTPPGGELVF 357
QY 366 ELWQNPNDHQRYAVKMFYQTMQDLRNAEKLKLNKNPAGIISVAVAGCE 414
Db 358 ERWRRLSDNSQWIVQSLVFTLQMRDKTPLSL-NTPPGEVKLTLAGCE 405

RESULT 3

US-09-291-931-2
; Sequence 2, Application US/09291931A
; Patent No. 6190897
; GENERAL INFORMATION:
; APPLICANT: Kretz, Keith
; TITLE OF INVENTION: NOVEL PHYTASE
; FILE REFERENCE: 09010/029003
; CURRENT APPLICATION NUMBER: US/09/291,931A
; CURRENT FILING DATE: 1999-04-13
; EARLIER APPLICATION NUMBER: 08/910,798
; EARLIER FILING DATE: 1997-08-13
; EARLIER APPLICATION NUMBER: 09/259,214

; EARLIER FILING DATE: 1999-03-01
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-291-931-2

Query Match 42.4%; Score 932.5; DB 3; Length 440;
Best Local Similarity 47.7%; Pred. No. 4.9e-92;
Matches 195; Conservative 61; Mismatches 138; Indels 15; Gaps 7;
QY 13 LVLMLSLAAITAPVA---AEPSCYTLERVVILSRHGVRSPTKOTQLMNDVTPDKWPMP 69
Db 5 LIPFLSLIPLTPQSAFAQSEPE-LKLESVVIVSRHGVRAPTKATQLMQDVTDPDWTWP 63
QY 70 VKAGYLPFGAELVTLMGGFYGYFRSLGLLA-AGCPAEGGVYAQADIDORTLTGQAF 128
Db 64 VLGLWLTFRGGLTAYLGHYQORLVADGLLAKKGCPSQGVIAIDVDETRTKTGEAFA 123
QY 129 DGVAPEGGLTVHQAADLKTDPLFHPVETGVCKLDNAQTDKAIEERLGGPLDVTVSRYAK 188
Db 124 AGLAPDCAITVHTQADTSSPDPLFNPKTGVCQLDNANVTDAILSRAGGSIAFTGHROT 183
QY 189 PFAQMGDVLPFAASPYCKSLQOQKTCDFAHFAANEVNVNKEGTVKLTSGPLALSSTLGE 248
Db 184 AFRELRLVLPQSNLCKREKQDECSLTQALPSELKVSAD--NVSLTGAVSLASMLTE 241
QY 249 IFLQNAQAMPEVAWQRLKGAENVVSLSLHNAQFNLMAKTPYIARHKGTPLLQQIDTAL 308
Db 242 IFLQQAQGMPEPGWGRITDQSHQNTLLSLHNAQFYLLQRTPEVARSRATPDLDLMAAL 301
QY 309 T---LQDAQOQKLPISAQNRVFLGGHDTNIAAGMLGADWQLPEQPDNTPPGGGLVF 365
Db 302 TPHPPOQAQYGVTLPTS---VLFIAGHDTNLANLGGALELNTLPGQPDNTPPGGELVF 357
QY 366 ELWQNPNDHQRYAVKMFYQTMQDLRNAEKLKLNKNPAGIISVAVAGCE 414
Db 358 ERWRRLSDNSQWIVQSLVFTLQMRDKTPLSL-NTPPGEVKLTLAGCE 405

RESULT 4

US-09-540-149A-1
; Sequence 1, Application US/09540149A
; Patent No. 6511699
; GENERAL INFORMATION:
; APPLICANT: Lei, Xingen
; TITLE OF INVENTION: ENZYMES WITH IMPROVED PHYTASE ACTIVITY
; FILE REFERENCE: 19603/2791
; CURRENT APPLICATION NUMBER: US/09/540,149A
; CURRENT FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 60/127,032
; PRIOR FILING DATE: 1999-03-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (433)
; OTHER INFORMATION: Xaa at position 433 in this sequence is unknown
US-09-540-149A-1

Query Match 42.2%; Score 928.5; DB 4; Length 433;
Best Local Similarity 47.4%; Pred. No. 1.3e-91;
Matches 194; Conservative 63; Mismatches 137; Indels 15; Gaps 7;
QY 13 LVLMLSLAAITAPVA---AEPSCYTLERVVILSRHGVRSPTKOTQLMNDVTPDKWPMP 69
Db 5 LIPFLSLIPLTPQSAFAQSEPE-LKLESVVIVSRHGVRAPTKATQLMQDVTDPDWTWP 63

[illegible]

RESIT.T 5

```

US-08-910-798-2
Sequence 2, Application US/08910798
Patent No. 5876997
GENERAL INFORMATION:
APPLICANT: KRETZ
TITLE OF INVENTION: NOVEL PHYTAL
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Fish & Richardson
STREET: 4225 Executive Square
CITY: La Jolla
STATE: California
COUNTRY: US
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-D
SOFTWARE: Patentin Release #1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/910
FILING DATE: August 13, 1997
CLASSIFICATION: 345
ATTORNEY/AGENT INFORMATION:
NAME: HAILL, PH.D., LISA A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 0901
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 423 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-910-798-2

```

Query Match 39.1%; Score 860; DB 2; Length 423;
Best Local Similarity 45.5%; Pred. No. 3.3e-84;
Matches 186; Conservative 58; Mismatches 133; Indels 32; Gaps 8;

QY 13 LVLMLSGLAAITAPVA--APPSTYLERVWILSRHGVRSPTKQTQLMNDVTPDKWPQW 69

Db 5 LIPFLSLILPTPOSFAQAEPE-LKLESVIVSRHGVRAPTKATQLMODVTPDAMPW 63

[illegible]

RESULT 6

```

US-09-489-039A-7512
; Sequence 7512, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breston et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 7512
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
; US-09-489-039A-7512

```

Query Match	22.6%	Score 498;	DB 4;	Length 421;
Best Local Similarity	30.5%;	Pred. NO. 5.5e-45;		
Matches 126:	Conservative	84;	Mismatches 165;	Indels 38;
				Gaps 12;

Qy	10	L S G L V M L S G L A A I T A P V A A E P S G Y T L R K W I L S H G V R S P - T K O T Q L M N D V T P F K W P Q W	68
D b	16	V T C A V M L S A G - - - A O A Q D K A A P E G Y C L Q O V L I M S R H N L A P L A N N G S V L E Q S T A K A W P Q W	72
Qy	69	P Y K A G Y L T P R G A E L V T M L M G F Y G D Y F R S L G L L A A G - C P A E G G V Y A A Q A D I D Q S T R L T G Q A F	127
D b	73	D V P G Q L T T K G G V L E Y V T M G H Y M E W L A Q O K U V T S G E C P E N A V A Y A N S L Q R T V A T A Q F F	132
Qy	128	L D G V A P G C G L T V H N O A D L K K T D P L F H P V E T G V C K L D N A C T D K A I B E R L G G P L D T V S Q R V A	187
D b	133	I T A F P G C G I P V H H G P O M G M T P T N P V I D D S P A P R E K A L O A M E K E R G - - - - - M Q L T	186
Qy	188	K X F A O M G D V L N A A S P Y C K S L O Q Q G K T C D F A H - - - F A A N E V N V N K E G T K V T L S G P L A L S	243
D b	187	E S Y K L L E T I D Y R N S P S C K - - - E K K V C S L S E G K D T F S A G Y Q Q - - - - - E P G V S G P L K V G	236
Qy	244	S T L G E I F L L O N A Q M P - - E V A N Q R L K G A E N W T V L L S L H N A Q N L M A K T Y P I A R H K T P L L	301
D b	237	N S L V D A F T L Q Y E G F P K O V A G E I A S D K O W A V L S K L K N G Y Q D S L T S V A V A Q N A K P L V	296
Qy	302	Q Q I D T A L T L Q L D A Q G K L P I S A Q N R V L F L G G H D T N I A T M L G - A D W L P E Q P D N T P P G	360
D b	297	K Y I D N A L V - - - G E G - - - - A S K A K Y T L V A G H D S N I A S L L T A L D P K Y O L P G O V E R T P I G	347

QY 361 GGLVFELWQNDHORYVAVKMFYQTMDOLENAEKLDLKNPAGIISVAVAGC 413
Db 348 GKLLFQWHDSDAGNRDLMKIEYVYQSTEQLEARNADALTQAPQR-VTIALNGC 399

RESULT 7
US-09-489-039A-13501
; Sequence 13501, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489, 039A
; PRIOR FILING DATE: 2000-01-27
; PRIOR FILING DATE: 2000-01-27
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 13501
; LENGTH: 522
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13501

Query Match 20.0%; Score 440.5; DB 4; Length 522;
Best Local Similarity 31.7%; Pred. No. 1.4e-38;
Matches 126; Conservative 62; Mismatches 178; Indels 31; Gaps 13;
QY 9 RLGLV-LMISGLAAITAPVAAPSGYTLERVVILSRHGVRSPTK-OTOLMNDVTPDKWP 66
Db 105 RHQGLRLFTACALPLLALQSAADWQLEKVVLSRHGIRPTAGNREAIATGRPT 164
QY 67 QWPKAGYLTPRGAEVLTLMGGFYGFYFRSLGLLAAGCAEAGGVYAQAIDORTRLTGOA 126
Db 165 EWTHTDGLTGHGVAAVNVKRAEGQHYRQLGLQACGTABSIYVRASPLQRTATAQA 224
QY 127 FLQGVAPCGGLTVHNOADLKTDPLEPHVETGVCKLDNAQTDKAIERLGGPLDVTVSQRY 186
Db 225 LVDAFPGCGVAHYVSG--DADPLFTQDKFAATQDTPARQLAAVKEAG---DLAQRQ 279
QY 187 AKPFAQMGVNLFAAPSYCKSLQOQKTCDFAHFAANEVNVNKEGKVTLSPLALSSLT 246
Db 280 A-----LAPTIQLLKQAVC---QADKPCPIFD-TPWQVEQSKG-KTTISG-LSVMANN 327
QY 247 GEIFLLQNAAMP--EVAWORLKAENWVLSLHNAQNLMAKTPYIARHKGTPLLQOI 304
Db 328 VETIRLQWSENPLPSQLANGKITCARQITALLPTEYDITNDVLYTAQKRGSVLLNAM 387
QY 305 DTALTLDAGQOKLPISAQNRVFLGGHDTNIAIAGMLGADWQLP-EQPDNTPPGGGL 363
Db 388 -----LD--GVKPEANPNVRWLLLVADHTNIAIVRTLMNFSWQLPGYSRGNIPPGSSL 438
QY 364 VFELWQNDHORYVAVKMFYQTMDOLENAEKLDLKN 400
Db 439 VLERWNAKSGERYLVYFOAQLDLDLRLQTPDAQH 475

RESULT 8
US-09-684-855-112
; Sequence 112, Application US/09684855
; Patent No. 6599735
; GENERAL INFORMATION:
; APPLICANT: F. Hoffmann-La Roche AG
; TITLE OF INVENTION: CONTINUOUS FERMENTATION PROCESS
; FILE REFERENCE: C38435/111692
; CURRENT APPLICATION NUMBER: US/09/684, 855
; PRIOR FILING DATE: 2000-10-06
; PRIOR FILING DATE: 2000-10-04
; PRIOR FILING DATE: 1999-10-11
; NUMBER OF SEQ ID NOS: 169
; SEQ ID NO 112
; LENGTH: 443
; TYPE: PRT
; ORGANISM: T. thermophilus
US-09-684-855-112

Query Match 5.2%; Score 115; DB 4; Length 443;
Best Local Similarity 22.1%; Pred. No. 0.0019;
Matches 87; Conservative 53; Mismatches 134; Indels 118; Gaps 18;
QY 31 PSGYTLERVVILSRHGVRSPTKQ-----TOLMNDVTPD-----KWPQWPKVAGY 74
Db 42 PQNCKITFVQLLSRHGARYPTSSKTELYSQLISRIQKTATAYKGYAFLKDYRYQLGAND 101
QY 75 LTPRGAEVLTLMGGFYGFYFRSLGLLAAGCAEAGGVYAQAIDORTRLTGOAFLDGVAPG 134
Db 102 LTPFGENQMIQLGKIFKNHYKSLARNV-----PPVRCSGSDRVIASGRFLFIEGFQSA 154
QY 135 CGLTVHNOADLKTDPLEPHV-----ETGVCKLDNAQTDKAIERLGGPLDVTVS 183
Db 155 KVLDPH--SDKHDAPPTINVIIIEGSPSYNNTLTGSCFV-----FEDSSG3--HDAQ 202
QY 184 QRYAKPFAQM-----GDVLNFAASPYCKSLQOQKTCDFAHFAANEVNVNKEGTK 233
Db 203 EKFAQFAPALLEKIKDHLPGVDLAVSDVPLMDL-----CPETLARNHTD----- 249
QY 234 VTLSGPAALSTLGEIFLQNAQMPVAVQRLKGAENWVLSLSLH--NAQFNLMAK--- 288
Db 250 -TLSPFCALST-----QSEWQ-----AYDYQSLGKYGNGGNGPLGPAQG 289
QY 289 ---TPYARHKGTPLLQOQIDTALTLDAGQOKLPISAQNRVFLGGHDTNIAIAGML 344
Db 290 VGFVNEILLIARHTSPV--QDYTVVNTLSDNPFATFLNATLYADF--SHDNTWTSIFAAL 345
QY 345 Q-----ADWQLPEQPD-----NTPFGGLVPELMQNDPNHORYVAV 380
Db 346 GLYNGTAKLSTTEIKSIEETDGSAAMTVFGRAYIEMMQCDDSDPEVVRV 397

RESULT 9
US-09-684-855-134
; Sequence 134, Application US/09684855
; Patent No. 6599735
; GENERAL INFORMATION:
; APPLICANT: F. Hoffmann-La Roche AG
; TITLE OF INVENTION: CONTINUOUS FERMENTATION PROCESS
; FILE REFERENCE: C38435/111692
; CURRENT APPLICATION NUMBER: US/09/684, 855
; PRIOR FILING DATE: 2000-10-06
; PRIOR FILING DATE: 2000-10-04
; PRIOR FILING DATE: 1999-10-11
; NUMBER OF SEQ ID NOS: 169
; SEQ ID NO 134
; LENGTH: 443
; TYPE: PRT
; ORGANISM: T. thermophilus
US-09-684-855-134

Query Match 5.2%; Score 115; DB 4; Length 443;
Best Local Similarity 22.1%; Pred. No. 0.0019;
Matches 87; Conservative 53; Mismatches 134; Indels 118; Gaps 18;
QY 31 PSGYTLERVVILSRHGVRSPTKQ-----TOLMNDVTPD-----KWPQWPKVAGY 74
Db 42 PQNCKITFVQLLSRHGARYPTSSKTELYSQLISRIQKTATAYKGYAFLKDYRYQLGAND 101
QY 75 LTPRGAEVLTLMGGFYGFYFRSLGLLAAGCAEAGGVYAQAIDORTRLTGOAFLDGVAPG 134
Db 102 LTPFGENQMIQLGKIFKNHYKSLARNV-----PPVRCSGSDRVIASGRFLFIEGFQSA 154

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 112
; LENGTH: 443
; TYPE: PRT
; ORGANISM: T. thermophilus
US-09-684-855-112

Query Match 5.2%; Score 115; DB 4; Length 443;
Best Local Similarity 21.1%; Pred. No. 0.0019;
Matches 87; Conservative 53; Mismatches 154; Indels 118; Gaps 18;
QY 31 PSGYTLERVVILSRHGVRSPTKQ-----TOLMNDVTPD-----KWPQWPKVAGY 74
Db 42 PQNCKITFVQLLSRHGARYPTSSKTELYSQLISRIQKTATAYKGYAFLKDYRYQLGAND 101
QY 75 LTPRGAEVLTLMGGFYGFYFRSLGLLAAGCAEAGGVYAQAIDORTRLTGOAFLDGVAPG 134
Db 102 LTPFGENQMIQLGKIFKNHYKSLARNV-----PPVRCSGSDRVIASGRFLFIEGFQSA 154
QY 135 CGLTVHNOADLKTDPLEPHV-----ETGVCKLDNAQTDKAIERLGGPLDVTVS 183
Db 155 KVLDPH--SDKHDAPPTINVIIIEGSPSYNNTLTGSCFV-----FEDSSG3--HDAQ 202
QY 184 QRYAKPFAQM-----GDVLNFAASPYCKSLQOQKTCDFAHFAANEVNVNKEGTK 233
Db 203 EKFAQFAPALLEKIKDHLPGVDLAVSDVPLMDL-----CPETLARNHTD----- 249
QY 234 VTLSGPAALSTLGEIFLQNAQMPVAVQRLKGAENWVLSLSLH--NAQFNLMAK--- 288
Db 250 -TLSPFCALST-----QSEWQ-----AYDYQSLGKYGNGGNGPLGPAQG 289
QY 289 ---TPYARHKGTPLLQOQIDTALTLDAGQOKLPISAQNRVFLGGHDTNIAIAGML 344
Db 290 VGFVNEILLIARHTSPV--QDYTVVNTLSDNPFATFLNATLYADF--SHDNTWTSIFAAL 345
QY 345 Q-----ADWQLPEQPD-----NTPFGGLVPELMQNDPNHORYVAV 380
Db 346 GLYNGTAKLSTTEIKSIEETDGSAAMTVFGRAYIEMMQCDDSDPEVVRV 397

RESULT 9
US-09-684-855-134
; Sequence 134, Application US/09684855
; Patent No. 6599735
; GENERAL INFORMATION:
; APPLICANT: F. Hoffmann-La Roche AG
; TITLE OF INVENTION: CONTINUOUS FERMENTATION PROCESS
; FILE REFERENCE: C38435/111692
; CURRENT APPLICATION NUMBER: US/09/684, 855
; PRIOR FILING DATE: 2000-10-06
; PRIOR FILING DATE: 2000-10-04
; PRIOR FILING DATE: 1999-10-11
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 134
; LENGTH: 443
; TYPE: PRT
; ORGANISM: T. thermophilus
US-09-684-855-134

Query Match 5.2%; Score 115; DB 4; Length 443;
Best Local Similarity 22.1%; Pred. No. 0.0019;
Matches 87; Conservative 53; Mismatches 134; Indels 118; Gaps 18;
QY 31 PSGYTLERVVILSRHGVRSPTKQ-----TOLMNDVTPD-----KWPQWPKVAGY 74
Db 42 PQNCKITFVQLLSRHGARYPTSSKTELYSQLISRIQKTATAYKGYAFLKDYRYQLGAND 101
QY 75 LTPRGAEVLTLMGGFYGFYFRSLGLLAAGCAEAGGVYAQAIDORTRLTGOAFLDGVAPG 134
Db 102 LTPFGENQMIQLGKIFKNHYKSLARNV-----PPVRCSGSDRVIASGRFLFIEGFQSA 154

QY 135 CGLTVHNOADLKKTDPLFHPV-----ETGVCKLDNAQTDKAEERLGGPLDTS 183
Db 155 KVLDPH--SDKHDAPPTINVIIEGSPSYNTLDTGSCP-----FEDSSGG--HDAQ 202
QY 184 QRYAKPFAQM-----GVLNFAASPYCKSLQOQKTCDFAHFAANEVNVNKEGTY 233
Db 203 EKFAQFAPALEIKIKDHLPGVDLAVSDVPYLMDL-----CPFETLARNHTD----- 249
QY 234 VTLSPGALSSITLGEIFLLQNAQAMPEVAVQRLKGAENWVLSLH--NAQFNLMAK--- 288
Db 250 -TLPFCALST-----QBEWQ-----AYDYOSLGKYGNGGNGPLGPAQ 289
QY 289 ----TPYIARHKGTPLLQOQIDTALTQLDAQOQKLPISAQNRVFLGQHDNTNIAAGML 344
Db 290 VGFVNELIARMTSPV--QDYTTVNHTLDSNPATFPLNATLYADF--SHDNTWTSIFAAL 345
QY 345 G-----ADWQLPEQPD-----NTPPGGLVFLWQPNHQRVAV 380
Db 346 GLYNGTAKLSTTEIKSIEETDGYSAANTVPFGGRAYIEMMQCDDSDPVRV 397

RESULT 10
US-09-684-855-157
; Sequence 157, Application US/09684855
; Patent No. 6599735
; GENERAL INFORMATION:
; APPLICANT: F. Hoffmann-La Roche AG
; TITLE OF INVENTION: CONTINUOUS FERMENTATION PROCESS
; FILE REFERENCE: C38435/111692
; CURRENT APPLICATION NUMBER: US/09/684,855
; CURRENT FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: EP 00121663.9
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: EP 99120289.6
; PRIOR FILING DATE: 1999-10-11
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 157
; LENGTH: 443
; TYPE: PRT
; ORGANISM: T. thermophilus
US-09-684-855-157

Query Match 5.2%; Score 115; DB 4; Length 443;
Best Local Similarity 21.1%; Pred. No. 0.0019;
Matches 87; Conservative 53; Mismatches 154; Indels 118; Gaps 18;

QY 31 PSGYTLERVILSRHGVRSPTKQ-----TQLMNDVTPD-----KWPQWPKVAGY 74
Db 42 PQNCKITFVQLLSRHGARYPTSSKTELYSLISRIQKTATAYKGYAFKDYRYQLGAND 101
QY 75 LTPRGAELVTLMGGFYGYFRSLGALLAACGPAEGVYQAADIDQRTLTGQAFDGVAPG 134
Db 102 LTPFGENQMIQLGIKFYNYHKSRLARNV-----PFVRCSGSDRVIASGLFIEGFQSA 154
QY 135 CGLTVHNOADLKKTDPLFHPV-----ETGVCKLDNAQTDKAEERLGGPLDTS 183
Db 155 KVLDPH--SDKHDAPPTINVIIEGSPSYNTLDTGSCP-----FEDSSGG--HDAQ 202
QY 184 QRYAKPFAQM-----GVLNFAASPYCKSLQOQKTCDFAHFAANEVNVNKEGTY 233
Db 203 EKFAQFAPALEIKIKDHLPGVDLAVSDVPYLMDL-----CPFETLARNHTD----- 249
QY 234 VTLSPGALSSITLGEIFLLQNAQAMPEVAVQRLKGAENWVLSLH--NAQFNLMAK--- 288
Db 250 -TLPFCALST-----QBEWQ-----AYDYOSLGKYGNGGNGPLGPAQ 289
QY 289 ----TPYIARHKGTPLLQOQIDTALTQLDAQOQKLPISAQNRVFLGQHDNTNIAAGML 344
Db 290 VGFVNELIARMTSPV--QDYTTVNHTLDSNPATFPLNATLYADF--SHDNTWTSIFAAL 345
QY 345 G-----ADWQLPEQPD-----NTPPGGLVFLWQPNHQRVAV 380

RESULT 11
US-09-044-718-9
; Sequence 9, Application US/09044718
; Patent No. 6391605
; GENERAL INFORMATION:
; APPLICANT: KOSTREWA, Dirk
; APPLICANT: PASANMONTES, Luis
; APPLICANT: TOMSCHY, Andrea
; APPLICANT: van LOON, Adolphus
; APPLICANT: VOGEL, Kurt
; APPLICANT: WYSS, Markus
; TITLE OF INVENTION: MODIFIED PHYTASES
; FILE REFERENCE: Modified Phytases
; CURRENT APPLICATION NUMBER: US/09/044,718
; CURRENT FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: EP 97810175.6
; PRIOR FILING DATE: 1997-03-25
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 450
; TYPE: PRT
; ORGANISM: Talaromyces thermophilus
US-09-044-718-9

Query Match 5.2%; Score 115; DB 4; Length 450;
Best Local Similarity 21.1%; Pred. No. 0.002;
Matches 87; Conservative 53; Mismatches 154; Indels 118; Gaps 18;

QY 31 PSGYTLERVILSRHGVRSPTKQ-----TQLMNDVTPD-----KWPQWPKVAGY 74
Db 49 PQNCKITFVQLLSRHGARYPTSSKTELYSLISRIQKTATAYKGYAFKDYRYQLGAND 108
QY 75 LTPRGAELVTLMGGFYGYFRSLGALLAACGPAEGVYQAADIDQRTLTGQAFDGVAPG 134
Db 109 LTPFGENQMIQLGIKFYNYHKSRLARNV-----PFVRCSGSDRVIASGLFIEGFQSA 161
QY 135 CGLTVHNOADLKKTDPLFHPV-----ETGVCKLDNAQTDKAEERLGGPLDTS 183
Db 162 KVLDPH--SDKHDAPPTINVIIEGSPSYNTLDTGSCP-----FEDSSGG--HDAQ 209
QY 184 QRYAKPFAQM-----GVLNFAASPYCKSLQOQKTCDFAHFAANEVNVNKEGTY 233
Db 210 EKFAQFAPALEIKIKDHLPGVDLAVSDVPYLMDL-----CPFETLARNHTD----- 256
QY 234 VTLSPGALSSITLGEIFLLQNAQAMPEVAVQRLKGAENWVLSLH--NAQFNLMAK--- 288
Db 257 -TLPFCALST-----QBEWQ-----AYDYOSLGKYGNGGNGPLGPAQ 296
QY 289 ----TPYIARHKGTPLLQOQIDTALTQLDAQOQKLPISAQNRVFLGQHDNTNIAAGML 344
Db 297 VGFVNELIARMTSPV--QDYTTVNHTLDSNPATFPLNATLYADF--SHDNTWTSIFAAL 352
QY 345 G-----ADWQLPEQPD-----NTPPGGLVFLWQPNHQRVAV 380
Db 353 GLYNGTAKLSTTEIKSIEETDGYSAANTVPFGGRAYIEMMQCDDSDPVRV 404

RESULT 12
US-08-868-435-31
; Sequence 31, Application US/08868435
; Patent No. 6291221
; GENERAL INFORMATION:
; APPLICANT: Van Loon, Adolphus
; APPLICANT: Mitchell, David
; TITLE OF INVENTION: POLYPEPTIDES WITH PHYTASE ACTIVITY
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann-La Roche Inc.

QY 135 CGLTVHNOADLKKTDPLFHPV-----ETGVCKLDNAQTDKAEERLGGPLDTS 183
Db 155 KVLDPH--SDKHDAPPTINVIIEGSPSYNTLDTGSCP-----FEDSSGG--HDAQ 202
QY 184 QRYAKPFAQM-----GVLNFAASPYCKSLQOQKTCDFAHFAANEVNVNKEGTY 233
Db 203 EKFAQFAPALEIKIKDHLPGVDLAVSDVPYLMDL-----CPFETLARNHTD----- 249
QY 234 VTLSPGALSSITLGEIFLLQNAQAMPEVAVQRLKGAENWVLSLH--NAQFNLMAK--- 288
Db 250 -TLPFCALST-----QBEWQ-----AYDYOSLGKYGNGGNGPLGPAQ 289
QY 289 ----TPYIARHKGTPLLQOQIDTALTQLDAQOQKLPISAQNRVFLGQHDNTNIAAGML 344
Db 290 VGFVNELIARMTSPV--QDYTTVNHTLDSNPATFPLNATLYADF--SHDNTWTSIFAAL 345
QY 345 G-----ADWQLPEQPD-----NTPPGGLVFLWQPNHQRVAV 380
Db 346 GLYNGTAKLSTTEIKSIEETDGYSAANTVPFGGRAYIEMMQCDDSDPVRV 397

RESULT 10
US-09-684-855-157
; Sequence 157, Application US/09684855
; Patent No. 6599735
; GENERAL INFORMATION:
; APPLICANT: F. Hoffmann-La Roche AG
; TITLE OF INVENTION: CONTINUOUS FERMENTATION PROCESS
; FILE REFERENCE: C38435/111692
; CURRENT APPLICATION NUMBER: US/09/684,855
; CURRENT FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: EP 00121663.9
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: EP 99120289.6
; PRIOR FILING DATE: 1999-10-11
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 157
; LENGTH: 443
; TYPE: PRT
; ORGANISM: T. thermophilus
US-09-684-855-157

Query Match 5.2%; Score 115; DB 4; Length 443;
Best Local Similarity 21.1%; Pred. No. 0.0019;
Matches 87; Conservative 53; Mismatches 154; Indels 118; Gaps 18;

QY 31 PSGYTLERVILSRHGVRSPTKQ-----TQLMNDVTPD-----KWPQWPKVAGY 74
Db 42 PQNCKITFVQLLSRHGARYPTSSKTELYSLISRIQKTATAYKGYAFKDYRYQLGAND 101
QY 75 LTPRGAELVTLMGGFYGYFRSLGALLAACGPAEGVYQAADIDQRTLTGQAFDGVAPG 134
Db 102 LTPFGENQMIQLGIKFYNYHKSRLARNV-----PFVRCSGSDRVIASGLFIEGFQSA 154
QY 135 CGLTVHNOADLKKTDPLFHPV-----ETGVCKLDNAQTDKAEERLGGPLDTS 183
Db 155 KVLDPH--SDKHDAPPTINVIIEGSPSYNTLDTGSCP-----FEDSSGG--HDAQ 202
QY 184 QRYAKPFAQM-----GVLNFAASPYCKSLQOQKTCDFAHFAANEVNVNKEGTY 233
Db 203 EKFAQFAPALEIKIKDHLPGVDLAVSDVPYLMDL-----CPFETLARNHTD----- 249
QY 234 VTLSPGALSSITLGEIFLLQNAQAMPEVAVQRLKGAENWVLSLH--NAQFNLMAK--- 288
Db 250 -TLPFCALST-----QBEWQ-----AYDYOSLGKYGNGGNGPLGPAQ 289
QY 289 ----TPYIARHKGTPLLQOQIDTALTQLDAQOQKLPISAQNRVFLGQHDNTNIAAGML 344
Db 290 VGFVNELIARMTSPV--QDYTTVNHTLDSNPATFPLNATLYADF--SHDNTWTSIFAAL 345
QY 345 G-----ADWQLPEQPD-----NTPPGGLVFLWQPNHQRVAV 380

```

; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; COUNTRY: United States of America
; ZIP: 07110
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/08/868,435
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/744,231
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Kass, Alan P
; REGISTRATION NUMBER: 32142
; REFERENCE/DOCKET NUMBER: Case Docket 9339
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201) 235-4205
; TELEFAX: (201) 235-2363
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 466 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 204
; OTHER INFORMATION: /note="potential N-glycosylation site"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 269
; OTHER INFORMATION: /note="potential N-glycosylation site"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 335
; OTHER INFORMATION: /note="potential N-glycosylation site"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 348
; OTHER INFORMATION: /note="potential N-glycosylation site"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 372
; OTHER INFORMATION: /note="potential N-glycosylation site"
; US-08-868-435-31

Query Match 5.28; Score 115; DB 3; Length 466;
Best Local Similarity 21.18; Pred. No. 0.0021;
Matches 87; Conservative 53; Mismatches 154; Indels 118; Gaps 18;

Qy 31 PSGYTLRVRVILSRHVSPTKQ-----TQIMNDVTPD-----KWQMPVKAGY 74
Db 65 PONCKITFVQLLSHGARYPTSSKTELYSQLISRIQKTATAYKYGYAFKDYRYQLGAND 124
Qy 75 LTPGAEVLTMGGFYGDYFRSLGLLAAGCPAGGVVAQADIDQRTLTGQAFLDGVAPG 134
Db 125 LTPFGENQMTQLGKIFKNHYKSLARNV-----PFVRCGSDRVIASGRLEIEGFQSA 177
Qy 135 CGLTVHVNQADLKTKDPLFHPV-----ETGVCKLDAQTDKAIERLGGPLDTYS 183
Db 178 KVLDPH--SDKHDPPTINIIIEGSPSYNNITDGTCPV-----FEDSSG--HDAQ 225
Qy 184 QRYAKPFAQM-----GDVLFNAASPYCKSLQQGKTCDFAHFAANEVNVNKEGTK 233
Db 226 EKPAKQAPALEIKDHPGVDLAVSDVPYMDL-----CPFETLARNHTD----- 272
Qy 234 VTLGGPLALSSTLGEIFLLQNAQAMPEVAQRKGAENWYLSLSLH--NAAQFNLAK--- 288

```

RESULT 13

```

US-08-744-231-31
; Sequence 31, Application US/08744231
; Patent No. 6358722
; GENERAL INFORMATION:
; APPLICANT: Van Loon, Adolphus
; APPLICANT: Mitchell, David
; TITLE OF INVENTION: POLYPEPTIDES WITH PHYTASE ACTIVITY
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; COUNTRY: United States of America
; ZIP: 07110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/744,231
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/424,757
; FILING DATE: 18-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Kass, Alan P
; REGISTRATION NUMBER: 32142
; REFERENCE/DOCKET NUMBER: Case Docket 9339
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201) 235-4205
; TELEFAX: (201) 235-2363
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 466 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 204
; OTHER INFORMATION: /note="potential N-glycosylation site"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 269
; OTHER INFORMATION: /note="potential N-glycosylation site"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 335
; OTHER INFORMATION: /note="potential N-glycosylation site"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 348
; OTHER INFORMATION: /note="potential N-glycosylation site"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 372
; OTHER INFORMATION: /note="potential N-glycosylation site"

```

US-08-744-231-31

Query Match 5.2%; Score 115; DB 4; Length 466;
Best Local Similarity 21.1%; Pred. No. 0.0021;
Matches 87; Conservative 53; Mismatches 154; Indels 118; Gaps 18;

QY 31 PSGYTLERVILSRHGVSPKQ-----TQLMNDVTPD-----KWPQMPVKAGY 74
DB 65 PONCKITFVQLLSRHGARYPTSSKTELYSQLISRIQKTATAYKGYAFELKDYRYQLGAND 124
QY 75 LTPRGAEVLTMGGFYGVYFSLGLLAAGCPAEGGVYQAADIDQRTLTGQAFLDGVAPG 134
DB 125 LTPFGENQMIQGIKFYKHYKSLARNV-----PFVRCSGSDRVIASGRFLIEGFQSA 177
QY 135 CGLTVHNOADLKKTDPLFHPV-----ETGVCKLDNAQTDKAIERLGGFLDVTVS 183
DB 178 KVLDPH--SDXHDAPPTINVIIIEGFSYNNLTDTGSCPV-----FEDSSGG--HDAQ 225
QY 184 QRYAKPFAQM-----GDVLFNFAASPYCKSLQOQKTCDFAHFAANEVNVNKEGTK 233
DB 226 EKFAQOFAPALEKIKDHLPGVDLAVSDVPYMLD-----CPFETLARNHTD----- 272
QY 234 VTLGSLFALSSTLGEIFLLQNAQAMPEVAWQRLKGAENWVLSLSH--NAQFNLMAK--- 288
DB 273 -TLSPPFCALST-----QBEWQ-----AYDYQSLGKYNGGNGPLGPAQ 312
QY 289 ----TPYIARHKGTPLLQIQIDTALTQLDAQOQKLPISAQNRVFLGCHDTNIANIAGML 344
DB 313 VGFVNELIARWTHSPV--QDYTVVNHTLDSNPATFPLNATLYADF--SHDNTMTSIFAAL 368
QY 345 G-----ADWQLEPEQD-----NTPPGGLVFLWQNPNDHQRVAV 380
DB 369 GLYNGTAKLSTTEIKSIEETDGYSAAWTVFGGRAYIEMWQCDSDSEPVVRV 420

RESULT 14

US-09-273-871A-13
; Sequence 13, Application US/09273871A
; Patent No. 6514495
; GENERAL INFORMATION:
; APPLICANT: Svendsen, Allan
; TITLE OF INVENTION: Phytase Variants
; FILE REFERENCE: 5618.500-US
; CURRENT APPLICATION NUMBER: US/09/273,871A
; CURRENT FILING DATE: 1999-03-22
; PRIOR APPLICATION NUMBER: PA 1998 00407
; PRIOR FILING DATE: 1998-03-23
; PRIOR APPLICATION NUMBER: PA 1998 00806
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: PA 1998 01176
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: PA 1999 00091
; PRIOR FILING DATE: 1999-01-22
; PRIOR APPLICATION NUMBER: 60/080,129
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/090,675
; PRIOR FILING DATE: 1998-06-25
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 466
; TYPE: PRT
; ORGANISM: Talaromyces thermophilus
US-09-273-871A-13

Query Match 5.2%; Score 115; DB 4; Length 466;
Best Local Similarity 21.1%; Pred. No. 0.0021;
Matches 87; Conservative 53; Mismatches 154; Indels 118; Gaps 18;

QY 31 PSGYTLERVILSRHGVSPKQ-----TQLMNDVTPD-----KWPQMPVKAGY 74
DB 65 PONCKITFVQLLSRHGARYPTSSKTELYSQLISRIQKTATAYKGYAFELKDYRYQLGAND 124

QY 75 LTPRGAEVLTMGGFYGVYFSLGLLAAGCPAEGGVYQAADIDQRTLTGQAFLDGVAPG 134
DB 125 LTPFGENQMIQGIKFYKHYKSLARNV-----PFVRCSGSDRVIASGRFLIEGFQSA 177
QY 135 CGLTVHNOADLKKTDPLFHPV-----ETGVCKLDNAQTDKAIERLGGFLDVTVS 183
DB 178 KVLDPH--SDXHDAPPTINVIIIEGFSYNNLTDTGSCPV-----FEDSSGG--HDAQ 225
QY 184 QRYAKPFAQM-----GDVLFNFAASPYCKSLQOQKTCDFAHFAANEVNVNKEGTK 233
DB 226 EKFAQOFAPALEKIKDHLPGVDLAVSDVPYMLD-----CPFETLARNHTD----- 272
QY 234 VTLGSLFALSSTLGEIFLLQNAQAMPEVAWQRLKGAENWVLSLSH--NAQFNLMAK--- 288
DB 273 -TLSPPFCALST-----QBEWQ-----AYDYQSLGKYNGGNGPLGPAQ 312
QY 289 ----TPYIARHKGTPLLQIQIDTALTQLDAQOQKLPISAQNRVFLGCHDTNIANIAGML 344
DB 313 VGFVNELIARWTHSPV--QDYTVVNHTLDSNPATFPLNATLYADF--SHDNTMTSIFAAL 368
QY 345 G-----ADWQLEPEQD-----NTPPGGLVFLWQNPNDHQRVAV 380
DB 369 GLYNGTAKLSTTEIKSIEETDGYSAAWTVFGGRAYIEMWQCDSDSEPVVRV 420

RESULT 15

US-09-252-991A-21854
; Sequence 21854, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21854
; LENGTH: 2291
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21854

Query Match 4.5%; Score 99; DB 4; Length 2291;
Best Local Similarity 22.7%; Pred. No. 1.7;
Matches 74; Conservative 38; Mismatches 116; Indels 98; Gaps 16;

QY 5 ENRVRLSGLVL-----MLSLAAITAPVA-----AEPGYTLERVV--ILSRHGVR 49
DB 1992 ENRRLLNGCVAVGNCACAEIRAEIDAGSTALNELVARQETANPGSDSDIAVGFLMGRNVVD 2051
QY 50 PTKQTQIMNDVTPDKV-----POMPVKAG--YLTPRGAEVLTMGGFYGVYFSLGLLAAGC 104
DB 2052 WTTAGQLHLEQTANLWNGNPGWQKEVGAYLDQTG-----FNPFGI---GV 2094
QY 105 PAEGGVYQAADIDQRTLTGQAFLDG-----VAPGCG--LTVHNOADLKKTDPLF 152
DB 2095 PAMGGAAG-----KVTAKALMNALKAGEIPKGEVAPGKANLPFGALADAEAGMPT 2146
QY 153 HPVE-----TGVCCKLDNAQTDKAIERLGGFLDVTVSQRYAKPFAQMGVDVLFNFAASPY 204
DB 2147 HPVKLAAKATGTAGKIKIEAGAIPIANEVRAGQGL-----SGLGYDVTHQTAS 2195
QY 205 CKSLQOQKTCDFAHFAANEVNV-----NKEGKVT-----LSGPLA-----LS 243
DB 2196 ARGIOQO--RTADLHVGLGSLDIVTTPKNDLPNTKIVRAIEKKNQAGGLVQADLPSTDM 2254
QY 244 STLGEIFLLQNAQAMPEVAWQRLKGA 269

Db 2255 SIAARMWGKTNQSIKTIFFQKPDGS 2280

Search completed: April 30, 2004, 12:43:16
Job time : 18.8293 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 30, 2004, 12:41:31 ; Search time 40.9756 Seconds
(without alignments)
2841.193 Million cell updates/sec

Title: US-10-021-723B-4
Perfect score: 2200
Sequence: 1 MSVLENRVRLSGVLMVLSGL.....NPAGIISVAVAGCENNGDDK 420

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1138130 seqs, 277199581 residues

Total number of hits satisfying chosen parameters: 1138120

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US05_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2200	100.0	420	14	US-10-021-723A-4
2	2193	99.7	441	14	US-10-021-723A-12
3	2183	99.2	441	12	US-10-282-122A-77792
4	2175	98.9	441	14	US-10-021-723A-2
5	948.5	43.1	436	14	US-10-156-660-2
6	943	42.9	410	14	US-10-021-723A-13
7	943	42.9	412	14	US-10-021-723A-15
8	943	42.9	412	14	US-10-334-672-1
9	943	42.9	412	14	US-10-334-671-1
10	939.5	42.7	432	9	US-09-866-379-8
11	939.5	42.7	432	12	US-10-282-122A-43351
12	939.5	42.7	432	14	US-10-156-660-4
13	932.5	42.4	440	9	US-09-777-568A-2
14	932.5	42.4	440	9	US-09-866-379-2
15	932.5	42.4	440	14	US-10-034-985-2

ALIGNMENTS

RESULT 1
US-10-021-723A-4
; Sequence 4, Application US/10021723A
; Publication No. US20030101476A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Mathur, Eric
; APPLICANT: Richardson, Toby
; APPLICANT: Robertson, Dan
; APPLICANT: Barton, Nelson
; TITLE OF INVENTION: Recombinant Phytases and Uses Thereof
; FILE REFERENCE: 112766.140 (DIV-016CIP)
; CURRENT APPLICATION NUMBER: US/10/021,723A
; CURRENT FILING DATE: 2002-10-23
; PRIOR APPLICATION NUMBER: US 60/255,090
; PRIOR FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 420
; TYPE: PRT
; ORGANISM: Versinia pestis
US-10-021-723A-4

Query Match	100.0%;	Score	2200;	DB	14;	Length	420;
Best Local Similarity	100.0%;	Pred. No.	4.1e-207;				
Matches	420;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
Qy	1	MSVLENRVRLSGVLMVLSGLAATAPVAAEPGVTLLRVVILSRHGVSRPTKOTQLMNDV	60				
Db	1	MSVLENRVRLSGVLMVLSGLAATAPVAAEPGVTLLRVVILSRHGVSRPTKOTQLMNDV	60				
Qy	61	TPQKWPQWPVKAGYLTIPRGAELVTLMGGFYGDYFRSLGLAAGCPAEGGYAQAIDQRT	120				
Db	61	TPQKWPQWPVKAGYLTIPRGAELVTLMGGFYGDYFRSLGLAAGCPAEGGYAQAIDQRT	120				
Qy	121	RLTQGAFLDGVAGCCGLTVNQADLKTDFLPFVETGVCKLNAQTDKAEERLGGPLD	180				
Db	121	RLTQGAFLDGVAGCCGLTVNQADLKTDFLPFVETGVCKLNAQTDKAEERLGGPLD	180				

Db 121 RLTAQAFLDGVAFCGGLTVHQAADLKKTDPLFHPVETGVCKLDNAQTDKAIEERLGGPLD 180
Qy 181 TVSORYAKPFAQMGDVNLNFAASPYCKSLQOQGTCTDFAHFAANEVNVNKEGTVKTLGSL 240
Db 181 TVSORYAKPFAQMGDVNLNFAASPYCKSLQOQGTCTDFAHFAANEVNVNKEGTVKTLGSL 240
Qy 241 ALSSTLGEIIFLLQNAQAMPEVAVQRLKGAENWVSLSLHNAQFNLMNAKTPYIARHKGTEL 300
Db 241 ALSSTLGEIIFLLQNAQAMPEVAVQRLKGAENWVSLSLHNAQFNLMNAKTPYIARHKGTEL 300
Qy 301 LQOQIDTALTQLDAQOQKLPISAQNRVFLFGGHDNTNIAIAGMLGADWQLPEQPDNTPPG 360
Db 301 LQOQIDTALTQLDAQOQKLPISAQNRVFLFGGHDNTNIAIAGMLGADWQLPEQPDNTPPG 360
Qy 361 GGLVFELWQPNHQRVAVVAVKMFYQTMQDLRNEAKLDLKNPAGIISVAVAGCENNGDDK 420
Db 361 GGLVFELWQPNHQRVAVVAVKMFYQTMQDLRNEAKLDLKNPAGIISVAVAGCENNGDDK 420

RESULT 2

US-10-021-723A-12
; Sequence 12, Application US/10021723A
; Publication No. US20030101476A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Mathur, Eric
; APPLICANT: Richardson, Toby
; APPLICANT: Robertson, Dan
; APPLICANT: Barton, Nelson
; TITLE OF INVENTION: Recombinant Phytases and Uses Thereof
; FILE REFERENCE: 112766.140 (DIV-016CIP)
; CURRENT APPLICATION NUMBER: US/10/021,723A
; CURRENT FILING DATE: 2002-10-23
; PRIOR FILING DATE: 2000-12-12
; PRIOR FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 441
; TYPE: PRT
; ORGANISM: Yersinia pestis
US-10-021-723A-12

Query Match 99.7%; Score 2193; DB 14; Length 441;
Best Local Similarity 99.8%; Pred. No. 2.1e-206;
Matches 419; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MSVLENRVLSGLVLMLSGLAAITAPVAAPSGYTLERVVILSRHGVRSPTKOTQLMNDV 60
Db 1 MSGLENRVLSGLVLMLSGLAAITAPVAAPSGYTLERVVILSRHGVRSPTKOTQLMNDV 60
Qy 61 TPDKWPQWPVKAAGYLTFRGAEVLTLMGFGYDYSRSLGILAAAGCPAEGGVYAQADIDQRT 120
Db 61 TPDKWPQWPVKAAGYLTFRGAEVLTLMGFGYDYSRSLGILAAAGCPAEGGVYAQADIDQRT 120
Qy 121 RLTAQAFLDGVAFCGGLTVHQAADLKKTDPLFHPVETGVCKLDNAQTDKAIEERLGGPLD 180
Db 121 RLTAQAFLDGVAFCGGLTVHQAADLKKTDPLFHPVETGVCKLDNAQTDKAIEERLGGPLD 180
Qy 181 TVSORYAKPFAQMGDVNLNFAASPYCKSLQOQGTCTDFAHFAANEVNVNKEGTVKTLGSL 240
Db 181 TVSORYAKPFAQMGDVNLNFAASPYCKSLQOQGTCTDFAHFAANEVNVNKEGTVKTLGSL 240
Qy 241 ALSSTLGEIIFLLQNAQAMPEVAVQRLKGAENWVSLSLHNAQFNLMNAKTPYIARHKGTEL 300
Db 241 ALSSTLGEIIFLLQNAQAMPEVAVQRLKGAENWVSLSLHNAQFNLMNAKTPYIARHKGTEL 300
Qy 301 LQOQIDTALTQLDAQOQKLPISAQNRVFLFGGHDNTNIAIAGMLGADWQLPEQPDNTPPG 360
Db 301 LQOQIDTALTQLDAQOQKLPISAQNRVFLFGGHDNTNIAIAGMLGADWQLPEQPDNTPPG 360
Qy 361 GGLVFELWQPNHQRVAVVAVKMFYQTMQDLRNEAKLDLKNPAGIISVAVAGCENNGDDK 420

Db 361 GGLVFELWQPNHQRVAVVAVKMFYQTMQDLRNEAKLDLKNPAGIISVAVAGCENNGDDK 420
RESULT 3
US-10-282-122A-77792
; Sequence 77792, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsuo
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 77792
; LENGTH: 441
; TYPE: PRT
; ORGANISM: Yersinia pestis
US-10-282-122A-77792

Query Match 99.2%; Score 2183; DB 12; Length 441;
Best Local Similarity 99.3%; Pred. No. 2.1e-205;
Matches 417; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MSVLENRVLSGLVLMLSGLAAITAPVAAPSGYTLERVVILSRHGVRSPTKOTQLMNDV 60
Db 1 MSVLENRVLSGLVLMLSGLAAITAPVAAPSGYTLERVVILSRHGVRSPTKOTQLMNDV 60
Qy 61 TPDKWPQWPVKAAGYLTFRGAEVLTLMGFGYDYSRSLGILAAAGCPAEGGVYAQADIDQRT 120
Db 61 TPDKWPQWPVKAAGYLTFRGAEVLTLMGFGYDYSRSLGILAAAGCPAEGGVYAQADIDQRT 120
Qy 121 RLTAQAFLDGVAFCGGLTVHQAADLKKTDPLFHPVETGVCKLDNAQTDKAIEERLGGPLD 180
Db 121 RLTAQAFLDGVAFCGGLTVHQAADLKKTDPLFHPVETGVCKLDNAQTDKAIEERLGGPLD 180
Qy 181 TVSORYAKPFAQMGDVNLNFAASPYCKSLQOQGTCTDFAHFAANEVNVNKEGTVKTLGSL 240
Db 181 TVSORYAKPFAQMGDVNLNFAASPYCKSLQOQGTCTDFAHFAANEVNVNKEGTVKTLGSL 240

QY 241 ALSSTLGEIFLQNAQAMPEVAWQRLKGAENWVLSLSLHNAQFNLMKTPYIARHKGTPPL 300
Db 241 ALSSTLGEIFLQNAQAMPEVAWQRLKGAENWVLSLSLHNAQFNLMKTPYIARHKGTPPL 300
QY 301 LQOITDALTQLDAGQKLPISAQNRVFLFGGHDNTNIANIAGMLGADWQLEQPDNTPPG 360
Db 301 LQOITDALTQLDAGQKLPISAQNRVFLFGGHDNTNIANIAGMLGADWQLEQPDNTPPG 360
QY 361 GGLVPELWQNPNDHORYVAVKMFYQTMQDLNRNAEKLDLKNPAGIISVAVAGCENNGDDK 420
Db 361 GGLVPELWQNPNDHORYVAVKMFYQTMQDLNRNAEKLDLKNPAGIISVAVAGCENNGDDK 420

RESULT 4

US-10-021-723A-2
; Sequence 2, Application US/10021723A
; Publication No. US20030101476A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Mathur, Eric
; APPLICANT: Richardson, Toby
; APPLICANT: Robertson, Dan
; APPLICANT: Barton, Nelson
; TITLE OF INVENTION: Recombinant Phycases and Uses Thereof
; FILE REFERENCE: 112766.140 (DIV-016CIP)
; CURRENT APPLICATION NUMBER: US/10/021,723A
; PRIOR FILING DATE: 2002-10-23
; PRIOR APPLICATION NUMBER: US 60/255,090
; PRIOR FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 441
; TYPE: PRT
; ORGANISM: Versinia pestis
; NAME/KEY: VARIANT
; LOCATION: 157, 163, 164, 174
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-021-723A-2

Query Match 98.9%; Score 2175; DB 14; Length 441;
Best Local Similarity 99.0%; Pred. No. 1.3e-204;
Matches 416; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MSVLENRVLSGLVLMLSGLAAITAPVAAPSGYTLERWILSRHGVSPKQTLQMDV 60
Db 1 MSVLENRVLSGLVLMLSGLAAITAPVAAPSGYTLERWILSRHGVSPKQTLQMDV 60
QY 61 TPDKWPQVKAAGYLTTPRGAELVTLMGFGYDYSRSLGGLAAGCPAEGGVYAADIDORT 120
Db 61 TPDKWPQVKAAGYLTTPRGAELVTLMGFGYDYSRSLGGLAAGCPAEGGVYAADIDORT 120
QY 121 RLTDGAFDGVAPGCGLTVRHQADLKKTDPLFHPVETGVCKLDNAQTDKAEERLGGPLD 180
Db 121 RLTDGAFDGVAPGCGLTVRHQADLKKTDPLFHPVETGVCKLDNAQTDKAEERLGGPLD 180
QY 181 TVSQRYAKPFQAGMDVINFAPASPYCKSLQOQKTCDFAHFAANEVNVNKEGTVTLGGPL 240
Db 181 TVSQRYAKPFQAGMDVINFAPASPYCKSLQOQKTCDFAHFAANEVNVNKEGTVTLGGPL 240
QY 241 ALSSTLGEIFLQNAQAMPEVAWQRLKGAENWVLSLSLHNAQFNLMKTPYIARHKGTPPL 300
Db 241 ALSSTLGEIFLQNAQAMPEVAWQRLKGAENWVLSLSLHNAQFNLMKTPYIARHKGTPPL 300
QY 301 LQOITDALTQLDAGQKLPISAQNRVFLFGGHDNTNIANIAGMLGADWQLEQPDNTPPG 360
Db 301 LQOITDALTQLDAGQKLPISAQNRVFLFGGHDNTNIANIAGMLGADWQLEQPDNTPPG 360
QY 361 GGLVPELWQNPNDHORYVAVKMFYQTMQDLNRNAEKLDLKNPAGIISVAVAGCENNGDDK 420
Db 361 GGLVPELWQNPNDHORYVAVKMFYQTMQDLNRNAEKLDLKNPAGIISVAVAGCENNGDDK 420

RESULT 5

US-10-156-660-2
; Sequence 2, Application US/10156660
; Publication No. US200301013958A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay M.
; APPLICANT: Kretz, Keith
; APPLICANT: Gray, Kevin A.
; APPLICANT: Barton, Nelson R.
; APPLICANT: Garrett, James B.
; APPLICANT: O'Donoghue, Eileen
; APPLICANT: Mathur, Eric J.
; TITLE OF INVENTION: PHYTASES, NUCLEIC ACIDS ENCODING THEM
; TITLE OF INVENTION: AND METHODS FOR MAKING AND USING THEM
; FILE REFERENCE: 09010-029007
; CURRENT APPLICATION NUMBER: US/10/156,660
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: US 09/866,379
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 09/580,515
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 09/318,528
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: US 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: US 09/259,214
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: US 08/910,798
; PRIOR FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 436
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: modified phytase enzyme
US-10-156-660-2

Query Match 43.1%; Score 948.5; DB 14; Length 436;
Best Local Similarity 48.2%; Pred. No. 4e-84;
Matches 197; Conservative 62; Mismatches 135; Indels 15; Gaps 7;

QY 13 LVLMLSGLAAITAPVA---ABPSGYTLERWILSRHGVSPKQTLQMDVTPDKVPQWP 69
Db 5 LIPFLSLLIPLTPQSAFAQSEPE-LKLESWIVSRHGVRAPTKATQLMQDVTDPDAWPTWP 63
QY 70 VKAGYLTTPRGAELVTLMGFGYDYSRSLGGLAAGCPAEGGVYAADIDORTLTGQAF 128
Db 64 VKUGELTPRGGELIATLGHYWRQRLVADGGLPKCGCPQSGQVATADVBERTKTGEAFA 123
QY 129 DGVAPCGGLTVHQAADLKKTDPLFHPVETGVCKLDNAQTDKAEERLGGPLDVTQRYAK 188
Db 124 AGLAPDCAITVHTQADTSSPDLFENLKTGVQCLDNANVTDAILERAGGSIADFTGHYQT 183
QY 189 PFAQMGDVINFAPASPYCKSLQOQKTCDFAHFAANEVNVNKEGTVTLGGPLALSITLGE 248
Db 184 AFRELERVINFPOSNCLKREKQDESCSLTQALPSELKVSAD--CVSLTGAVSLASMLTE 241
QY 249 IFLLQNAQAMPEVAWQRLKGAENWVLSLSLHNAQFNLMKTPYIARHKGTPPLIQQIDTAL 308
Db 242 IFLLQNAQAMPEVPGWGRITDSHQWNTLLSHNAQFDLLQRTPEVARSRTATLLDLIKTAL 301
QY 309 T----LQIDAGQKLPISAQNRVFLFGGHDNTNIANIAGMLGADWQLEQPDNTPPGGLVFP 365
Db 302 TPHPKQKQYGVTLPTS----VLFAGHDNTNLANLGGALELWNTLPGQPDNTPPGGELVFP 357
QY 366 ELWNPNDHORYVAVKMFYQTMQDLNRNAEKLDLKNPAGIISVAVAGCE 414
Db 358 ERWRERLSDNSQWISVLEFQTLQQRDKTFLSL-NTFPGEVKLTLAGCE 405

```

; CURRENT FILING DATE: 2002-10-23
; PRIOR APPLICATION NUMBER: US 60/255,090
; PRIOR FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 410
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: amino acid sequence of a phytase enzyme with
; OTHER INFORMATION: predicted glycosylation sites
US-10-021-723A-15

Query Match 42.9%; Score 943; DB 14; Length 410;
Best Local Similarity 49.0%; Pred. No. 1.3e-83;
Matches 191; Conservative 60; Mismatches 127; Indels 12; Gaps 6;

QY 29 AEPGTYLIERVILSRHGVSPTKQTLMDNDVTPDKWPQMPVKGAYLTPRGAEVLTMGG 88
Db 2 SEPE-LKLESVIVSRHGVRAPTKATQMQDVTDPDAMPTWFKLGLTTPRGELIAYLGH 60
QY 89 FYGDYFRLSGLL-AAGCPAEGGVYAQADIDORTLTQAFLDGVAPCGGLTVHNOADLKK 147
Db 61 YWRQLVADGLLPKCGCPQSGQVAIIADVDERTRKTGEAFAAGLAPDCAITVHTQADTSS 120
QY 148 TDPLFHPVETGVCKLDNAQTDKAEERLGGPLDVTVSQRYAKPFAQMGDVILNFAASPYCKS 207
Db 121 PDPLFNLKTVGCQDLNANVTDAILERAGGSIADFTGHYQTAFFELERVLNFPQSNLCLK 180
QY 208 LQOQKTCDFAHFAANEVNNKGTKTLSPALSTLGEIFLLQNAQAMPEVAVQRLK 267
Db 181 REKQDECSLTQALPSELKVSAD--CVSLTGAVSLASMLTEIFLLQOQMPPEPGWGRIT 238
QY 268 GAENWVLSLHNAQFNLMAKTPYIARHKGTPLLQQIDTALT---LQLDAQGQKLPISAQ 324
Db 239 DSHQWNLLSLHNAQFLLQRTPEVARSRAFPDLDLIKTALTPHPKQAYGVTLPST-- 296
QY 325 NRVLFLGHDNTNANIAGMLGADWOLPEQDNTPPGGLVPELWQNDHQRVAVTMFY 384
Db 297 --VLFIAGHDTNLANLGALLENWTLPGQDNTPPGGLVPELWQNDHQRVAVTMFY 354
QY 385 QTMQLRNAEKLDLKNPAGIISVAVAGCE 414
Db 355 QTLQMRDKTPLSL-NTPPGEVKTLAGCE 383

RESULT 8
US-10-334-672-1
; Sequence 1, Application US/10334672
; Publication No. US20030157646A1
; GENERAL INFORMATION:
; APPLICANT: Lananan, Mike
; APPLICANT: Koepf, Edward
; APPLICANT: Kretz, Keith
; TITLE OF INVENTION: Microbially-Expressed Thermotolerant Phytase For Animal Feed
; FILE REFERENCE: SYNG-P01-001
; CURRENT APPLICATION NUMBER: US/10/334,672
; CURRENT FILING DATE: 2002-12-30
; PRIOR APPLICATION NUMBER: 60/344,523
; PRIOR FILING DATE: 2001-12-28
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 412
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: No. US20030157646A19X Phytase
US-10-334-672-1

Query Match 42.9%; Score 943; DB 14; Length 412;
Best Local Similarity 49.0%; Pred. No. 1.3e-83;

```

```

; CURRENT FILING DATE: 2002-10-23
; PRIOR APPLICATION NUMBER: US 60/255,090
; PRIOR FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 410
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: amino acid sequence of a phytase enzyme with
; OTHER INFORMATION: predicted glycosylation sites
US-10-021-723A-13

Query Match 42.9%; Score 943; DB 14; Length 410;
Best Local Similarity 49.0%; Pred. No. 1.3e-83;
Matches 191; Conservative 60; Mismatches 127; Indels 12; Gaps 6;

QY 29 AEPGTYLIERVILSRHGVSPTKQTLMDNDVTPDKWPQMPVKGAYLTPRGAEVLTMGG 88
Db 2 SEPE-LKLESVIVSRHGVRAPTKATQMQDVTDPDAMPTWFKLGLTTPRGELIAYLGH 60
QY 89 FYGDYFRLSGLL-AAGCPAEGGVYAQADIDORTLTQAFLDGVAPCGGLTVHNOADLKK 147
Db 61 YWRQLVADGLLPKCGCPQSGQVAIIADVDERTRKTGEAFAAGLAPDCAITVHTQADTSS 120
QY 148 TDPLFHPVETGVCKLDNAQTDKAEERLGGPLDVTVSQRYAKPFAQMGDVILNFAASPYCKS 207
Db 121 PDPLFNLKTVGCQDLNANVTDAILERAGGSIADFTGHYQTAFFELERVLNFPQSNLCLK 180
QY 208 LQOQKTCDFAHFAANEVNNKGTKTLSPALSTLGEIFLLQNAQAMPEVAVQRLK 267
Db 181 REKQDECSLTQALPSELKVSAD--CVSLTGAVSLASMLTEIFLLQOQMPPEPGWGRIT 238
QY 268 GAENWVLSLHNAQFNLMAKTPYIARHKGTPLLQQIDTALT---LQLDAQGQKLPISAQ 324
Db 239 DSHQWNLLSLHNAQFLLQRTPEVARSRAFPDLDLIKTALTPHPKQAYGVTLPST-- 296
QY 325 NRVLFLGHDNTNANIAGMLGADWOLPEQDNTPPGGLVPELWQNDHQRVAVTMFY 384
Db 297 --VLFIAGHDTNLANLGALLENWTLPGQDNTPPGGLVPELWQNDHQRVAVTMFY 354
QY 385 QTMQLRNAEKLDLKNPAGIISVAVAGCE 414
Db 355 QTLQMRDKTPLSL-NTPPGEVKTLAGCE 383

RESULT 7
US-10-021-723A-15
; Sequence 15, Application US/10021723A
; Publication No. US20030101476A1
; GENERAL INFORMATION:
; APPLICANT: Mathur, Eric
; APPLICANT: Richardson, Toby
; APPLICANT: Robertson, Dan
; APPLICANT: Barton, Nelson
; TITLE OF INVENTION: Recombinant Phytases and Uses Thereof
; FILE REFERENCE: 112766.140 (DIV-016CIP)
; CURRENT APPLICATION NUMBER: US/10/021,723A
; CURRENT FILING DATE: 2002-10-23
; PRIOR APPLICATION NUMBER: US 60/255,090
; PRIOR FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 410
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Phytases and Uses Thereof
; FILE REFERENCE: 112766.140 (DIV-016CIP)
; CURRENT APPLICATION NUMBER: US/10/021,723A

```


QY 309 T---LQDAQOGKLPISQNRVFLFGHDTNINIAAGMLGADWOLPEQDNTTPGGGLVF 365
Db 302 TPHPKQKQAGVTLPTS-----VLFIAHDTNLANLGALLENWTLFGQDNTTPGGGLVF 357
QY 366 ELWQNDPNHORYVAVKQFVTQMDQLRNAEKLDKNNPAGIISVAVAGE 414
Db 358 ERWRRLSDNSQWITQVSLVFTQMQMRDKTFLSL-NTPPGEVKLTLAGCE 405

RESULT 11

US-10-282-122A-43351
; Sequence 43351, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELIURA 0343
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 43351
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Escherichia coli

US-10-282-122A-43351
Query Match 42.7%; Score 939.5; DB 12; Length 432;
Best Local Similarity 47.9%; Pred. No. 3e-83;
Matches 196; Conservative 61; Mismatches 137; Indels 15; Gaps 7;
QY 13 LVLMISGLAATAPVA---AEPSCYTLERVILSHGVRSPTKQTLMDVTPDKWPWP 69
Db 5 LIPFLSLIPLTPQSAFAQSEPE-LKLESVIVSRHGVRAPTKATQMLQMDVTPDAMP 63
QY 70 VKAGYLTFRGAEVLTMGMFGFYGYFRSLGLLA-AGCPAEGGVYAQADIDQRTLTGOAFL 128
Db 64 VKLGLWLTFRGELIAYLGHVQRLVADGLLAKKGCPSQGVAIADVDERTKTEGAFA 123
QY 129 DGVAPEGGLTVHQAADLKKTDPLPHVETGVCKLDNAQTDKAEERLGGPLDVTVSQYAK 188
Db 124 AGLADPCAITVHTQADTSSPDLFNPLKTVGCQLDNANVTDAILSRAGGSIAPTGHRQT 193
QY 189 PFAQMGDVINFAASPYCKSLQOQKTCDFAHFAANEVNVNKGKTKVLSGFLSSTLGE 248

Db 124 AGLADPCAITVHTQADTSSPDLFNPLKTVGCQLDNANVTDAILSRAGGSIAPTGHRQT 193
QY 189 PFAQMGDVINFAASPYCKSLQOQKTCDFAHFAANEVNVNKGKTKVLSGFLSSTLGE 248
Db 184 APELERVILFPQSNLCLKREKQDESCSLTQALPSLKVSAAD--NVSLTGAIVSLAMLTE 241
QY 249 IFLLQNAQAMPEVAMORUKGAENVVSLLSLHNAQFNIAWAKTPYIARHKGTPLLQOQIDTAL 308
Db 242 IFLLQQAQGMPEPGWGRITDSHQNTLLSLHNAQFYLLQRTPEVARSRATFLDLIKTAL 301
QY 309 T---LQDAQOGKLPISQNRVFLFGHDTNINIAAGMLGADWOLPEQDNTTPGGGLVF 365
Db 302 TPHPKQKQAGVTLPTS-----VLFIAHDTNLANLGALLENWTLFGQDNTTPGGGLVF 357
QY 366 ELWQNDPNHORYVAVKQFVTQMDQLRNAEKLDKNNPAGIISVAVAGE 414
Db 358 ERWRRLSDNSQWITQVSLVFTQMQMRDKTFLSL-NTPPGEVKLTLAGCE 405

RESULT 12

US-10-156-660-4
; Sequence 4, Application US/10156660
; Publication No. US20030103958A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay M.
; APPLICANT: Kretz, Keith
; APPLICANT: Gray, Kevin A.
; APPLICANT: Barton, Nelson R.
; APPLICANT: Garrett, James B.
; APPLICANT: O'Donoghue, Eileen
; APPLICANT: Mathur, Eric J.
; TITLE OF INVENTION: PHYTASES, NUCLEIC ACIDS ENCODING THEM
; TITLE OF INVENTION: AND METHODS FOR MAKING AND USING THEM
; FILE REFERENCE: 09010-029007
; CURRENT APPLICATION NUMBER: US/10/156,660
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: US 09/866,379
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 09/580,515
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 09/318,528
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: US 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: US 09/259,214
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: US 08/910,798
; PRIOR FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-156-660-4

Query Match 42.7%; Score 939.5; DB 14; Length 432;
Best Local Similarity 47.9%; Pred. No. 3e-83;
Matches 196; Conservative 61; Mismatches 137; Indels 15; Gaps 7;
QY 13 LVLMISGLAATAPVA---AEPSCYTLERVILSHGVRSPTKQTLMDVTPDKWPWP 69
Db 5 LIPFLSLIPLTPQSAFAQSEPE-LKLESVIVSRHGVRAPTKATQMLQMDVTPDAMP 63
QY 70 VKAGYLTFRGAEVLTMGMFGFYGYFRSLGLLA-AGCPAEGGVYAQADIDQRTLTGOAFL 128
Db 64 VKLGLWLTFRGELIAYLGHVQRLVADGLLAKKGCPSQGVAIADVDERTKTEGAFA 123
QY 129 DGVAPEGGLTVHQAADLKKTDPLPHVETGVCKLDNAQTDKAEERLGGPLDVTVSQYAK 188
Db 124 AGLADPCAITVHTQADTSSPDLFNPLKTVGCQLDNANVTDAILSRAGGSIAPTGHRQT 193
QY 189 PFAQMGDVINFAASPYCKSLQOQKTCDFAHFAANEVNVNKGKTKVLSGFLSSTLGE 248

Db 184 AFRELVNFPQSNLCKREKQDESCSLTQALPSELKVSAD--NVSUTGAVSLASMLTE 241
Qy 249 IFLLQNAQAMPEVAVQRLKGAENWVLSLSLHNAQFNLMKTPYIARHKGTPILLQOQIDTAL 308
Db 242 IFLLQNAQAMPEVAVQRLKGAENWVLSLSLHNAQFNLMKTPYIARHKGTPILLQOQIDTAL 301
Qy 309 T---LQDAQOGKLPISAGNRVFLFGGHDNTNIAAGMLGADWQLPEQPDNTPPGGGLVF 365
Db 302 TPHPKQAYGVTLPTS-----VLFIAGHDNTNLANLGGALELAWLTPGQDNTPPGGGLVF 357
Qy 366 ELWQNPDPHQRVYAVKMFYQTMQDLQRLNAEKLDLKNPAGIISVAVAGCE 414
Db 358 ERWRRLSDNSQWISQVSLVFQTLQOMDKTPLSL-NTPPGEVKLTLAGCE 405

RESULT 13
US-09-777-566A-2
; Sequence 2, Application US/09777566A
; Patent No. US20010055788A1
; GENERAL INFORMATION:
; APPLICANT: DIVERSA CORPORATION
; APPLICANT: SHORT, Jay
; APPLICANT: KRETZ, Keith
; APPLICANT: BARTON, Nelson
; APPLICANT: GARRETT, James
; APPLICANT: O'DONOGHUE, Eileen
; TITLE OF INVENTION: RECOMBINANT BACTERIAL PHYTASES AND USES THEREOF
; FILE REFERENCE: DIVER1370-6
; CURRENT APPLICATION NUMBER: US/09/777,566A
; PRIOR FILING DATE: 2001-06-11
; PRIOR APPLICATION NUMBER: US 09/318,528
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: US 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: US 09/259,214
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: US 08/910,798
; PRIOR FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1323)
; OTHER INFORMATION: n is any nucleotide
US-09-777-566A-2

Query Match 42.4%; Score 932.5; DB 9; Length 440;
Best Local Similarity 47.7%; Pred. No. 1.5e-82;
Matches 195; Conservative 61; Mismatches 138; Indels 15; Gaps 7;

Qy 13 LVLMLSGLAATAPVA---AEPGTYLTVLHSHVSRPTKQTLQMDVTPDKWPWP 69
Db 5 LIPFLSLIPLTPQSAFAQSEPE-LKLESVIVSRHGVRAPTKATQMLQMDVTPDAMPWP 63
Qy 70 VKAGYLTFRGAEVLTMGGFYGDYFRSLGLLA-ACPAEGGVYAAQDIDQRTLTGQAF 128
Db 64 VKLGWLTFRGAEVLTMGGFYGDYFRSLGLLA-ACPAEGGVYAAQDIDQRTLTGQAF 123
Qy 129 DGVAPCGGLTVHQAADLKKTDPLFHPVETGVCKLDNAQTKAIEERLGGPLDTPVSQRYAK 188
Db 124 AGLAPDCAITVHTQADTSSPDPLFNPLKTVGVCQLDNANVTDAISRAGGSIAFTGHRQT 183
Qy 189 PFAQMGDLVNFPAASPYCKSLQOQKTCDFAHFAANVAVNKEGTVKTLGSLPLASSTIGE 248
Db 184 AFRELVNFPQSNLCKREKQDESCSLTQALPSELKVSAD--NVSUTGAVSLASMLTE 241
Qy 249 IFLLQNAQAMPEVAVQRLKGAENWVLSLSLHNAQFNLMKTPYIARHKGTPILLQOQIDTAL 308
Db 242 IFLLQNAQAMPEVAVQRLKGAENWVLSLSLHNAQFNLMKTPYIARHKGTPILLQOQIDTAL 301
Qy 309 T---LQDAQOGKLPISAGNRVFLFGGHDNTNIAAGMLGADWQLPEQPDNTPPGGGLVF 365
Db 242 IFLLQNAQAMPEVAVQRLKGAENWVLSLSLHNAQFNLMKTPYIARHKGTPILLQOQIDTAL 301
Qy 309 T---LQDAQOGKLPISAGNRVFLFGGHDNTNIAAGMLGADWQLPEQPDNTPPGGGLVF 365

Db 302 TPHPKQAYGVTLPTS-----VLFIAGHDNTNLANLGGALELAWLTPGQDNTPPGGGLVF 357
Qy 366 ELWQNPDPHQRVYAVKMFYQTMQDLQRLNAEKLDLKNPAGIISVAVAGCE 414
Db 358 ERWRRLSDNSQWISQVSLVFQTLQOMDKTPLSL-NTPPGEVKLTLAGCE 405

RESULT 14
US-09-866-379-2
; Sequence 2, Application US/09866379
; Patent No. US20020136754A1
; GENERAL INFORMATION:
; APPLICANT: DIVERSA CORPORATION
; APPLICANT: SHORT, Jay
; APPLICANT: KRETZ, Keith
; APPLICANT: GRAY, Kevin
; APPLICANT: BARTON, Nelson
; APPLICANT: GARRETT, James
; APPLICANT: O'DONOGHUE, Eileen
; TITLE OF INVENTION: RECOMBINANT BACTERIAL PHYTASES AND USES THEREOF
; FILE REFERENCE: DIVER1370-7
; CURRENT APPLICATION NUMBER: US/09/866,379
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 09/580,515
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 09/318,528
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: US 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: US 09/259,214
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: US 08/910,798
; PRIOR FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 2
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-866-379-2

Query Match 42.4%; Score 932.5; DB 9; Length 440;
Best Local Similarity 47.7%; Pred. No. 1.5e-82;
Matches 195; Conservative 61; Mismatches 138; Indels 15; Gaps 7;

Qy 13 LVLMLSGLAATAPVA---AEPGTYLTVLHSHVSRPTKQTLQMDVTPDKWPWP 69
Db 5 LIPFLSLIPLTPQSAFAQSEPE-LKLESVIVSRHGVRAPTKATQMLQMDVTPDAMPWP 63
Qy 70 VKAGYLTFRGAEVLTMGGFYGDYFRSLGLLA-ACPAEGGVYAAQDIDQRTLTGQAF 128
Db 64 VKLGWLTFRGAEVLTMGGFYGDYFRSLGLLA-ACPAEGGVYAAQDIDQRTLTGQAF 123
Qy 129 DGVAPCGGLTVHQAADLKKTDPLFHPVETGVCKLDNAQTKAIEERLGGPLDTPVSQRYAK 188
Db 124 AGLAPDCAITVHTQADTSSPDPLFNPLKTVGVCQLDNANVTDAISRAGGSIAFTGHRQT 183
Qy 189 PFAQMGDLVNFPAASPYCKSLQOQKTCDFAHFAANVAVNKEGTVKTLGSLPLASSTIGE 248
Db 184 AFRELVNFPQSNLCKREKQDESCSLTQALPSELKVSAD--NVSUTGAVSLASMLTE 241
Qy 249 IFLLQNAQAMPEVAVQRLKGAENWVLSLSLHNAQFNLMKTPYIARHKGTPILLQOQIDTAL 308
Db 242 IFLLQNAQAMPEVAVQRLKGAENWVLSLSLHNAQFNLMKTPYIARHKGTPILLQOQIDTAL 301
Qy 309 T---LQDAQOGKLPISAGNRVFLFGGHDNTNIAAGMLGADWQLPEQPDNTPPGGGLVF 365
Db 302 TPHPKQAYGVTLPTS-----VLFIAGHDNTNLANLGGALELAWLTPGQDNTPPGGGLVF 357
Qy 366 ELWQNPDPHQRVYAVKMFYQTMQDLQRLNAEKLDLKNPAGIISVAVAGCE 414
Db 358 ERWRRLSDNSQWISQVSLVFQTLQOMDKTPLSL-NTPPGEVKLTLAGCE 405

```

RESULT 15
US-10-034-985-2
; Sequence 2, Application US/10034985
; Publication No. US20030049815A1
; GENERAL INFORMATION:
; APPLICANT: Kretz, Keith
; TITLE OF INVENTION: NOVEL PHYTASE
; FILE REFERENCE: 09010/029003
; CURRENT APPLICATION NUMBER: US/10/034,985
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: US/09/580,515
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: 08/910,798
; PRIOR FILING DATE: 1997-08-13
; PRIOR APPLICATION NUMBER: 09/259,214
; PRIOR FILING DATE: 1999-03-01
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-034-985-2

Query Match      42.4%; Score 932.5; DB 14; Length 440;
Best Local Similarity 47.7%; Pred. No. 1.5e-82;
Matches 195; Conservative 61; Mismatches 138; Indels 15; Gaps 7;

Qy 13 LVLMISGLAATAPV---APPSGYTLERVVILSRHGVRSPTKTOLMNDVTDKWPWP 69
Db 5 LIPFLSLIPITPOSAFAQSEPE-LKLESVVIVSRHGVRAPTKATQLMQDVTPDANPTWP 63
Qy 70 VKAGYLPFGAELVTLMGFGFYGYRSIGLIA-AGCPAEGGVAAQADIDQTRLTGQAF 128
Db 64 VKLGLWLPFGGELIAYLGHYQORQLVADGLLAKKGCPSQGVALLADVDERTRKTGEAFA 123
Qy 129 DGVAPGCGLTVHNOADLKKTDPLFHPVETGVCKLDNAQTDKAIERLGGPLDVTVSQRYAK 188
Db 124 AGLAPDCAITVHTQADTSSPDPLFNPLKTVGCQLDNNANVTDAILSRAGGSIADFTGHRQT 183
Qy 189 PFAQMGDVLFNAASPYCKSLQQCGKTCDFAHFAANEVNVNKEGTVKTLSPGLALSSTLGE 248
Db 184 AFRELERVLPQSNCLKREKQDESCSLTQALPSELKVSAD--NVSLTGAVSLASMLTE 241
Qy 249 IFLQNAQMEVAVQRLKGAENWVLSLHNAQFNMAKTPYIARHKGTPLQOQIDTAL 308
Db 242 IFLQQAQCMPEPGWGRITDSHOWNTLLSHNAQFYLLIQTPEVARSRAPEPLDLINAAAL 301
Qy 309 T---LQLDAQGCKLPISAQNEVLFLGGHDTNIANIAGMLGADWOLPEQPDNTPPGGGLVF 365
Db 302 TPHPKQAYGVTLPTS----VLFAGHDTNLANLGGALELNTWLPQPDNTPPGGELVF 357
Qy 366 ELWQNPDNHQRVAVKMFYQTMQDLRNAEKLDLKXNPAGIISVAVAGCE 414
Db 358 ERWRRLSDNSQWIQVSLVFQTLQQRDKTPLSL-NTPPGEVKLTLAGCE 405

```

Search completed: April 30, 2004, 12:55:02
Job time : 42.9756 secs

proteinase IV [imp
histidine acid pho
dynsin heavy chain
hypothetical prote
probable hemagglut
45K immediate earl
ABC-type transport
gene 16 protein -
hypothetical prote
O-antigen biosynth
B. subtilis Cggr h
adenylouccinate 1
hypothetical prote
probable autotrans
dynsin gamma heavy
phosphogluconate d

30 100 4.5 609 2 AF2379
31 99.5 4.5 465 2 JE0369
32 99.5 4.5 4367 1 B54802
33 99 4.5 440 2 T32457
34 99 4.5 3535 2 E83641
35 98.5 4.5 411 1 EDBEM4
36 98.5 4.5 586 2 F64186
37 98.5 4.5 609 1 A42537
38 98 4.5 436 2 T16638
39 97.5 4.4 348 2 AD1382
40 97.5 4.4 1275 2 T18556
41 96.5 4.4 467 2 T04540
42 96.5 4.4 527 2 T33175
43 95 4.3 3705 2 AD0123
44 95 4.3 4485 2 T08044
45 94.5 4.3 625 2 AC0309

ALIGNMENTS

RESULT 1
AC0201
acid phosphatase (EC 3.1.3.2) [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001
C:Accession: AC0201
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.M.; Holden, M.T.G.; Prentice, M.;
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, S.; Whitehead, S.; Barrrell
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AC0201
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-441 <XUR>
A:Cross-references: GB:AL590842; PIDN:CAC90470.1; PID:gl5979685; GSPDB:GN00175
C:Genetics:
A:Gene: YPO1648
C:Keywords: phosphoric monoester hydrolase

Query Match 99.2%; Score 2183; DB 2; Length 441;
Best Local Similarity 99.3%; Pred. No. 1.2e-168;
Matches 417; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 1 MSVLENRVRLSGVLMLSLGLAATAPVAEPSPGYTLERVVILSRHGVRSPKTKQLMNDV 60
Db 1 MSVLENRVRLSGVLMLSLGLAATAPVAEPSPGYTLERVVILSRHGVRSPKTKQLMNDV 60
Qy 61 TPDKWPMPVKAGYLTTPRGAEVLTLMGGFTYDIFRSLGLLAAGCPAEGGYAQAADIDQRT 120
Db 61 TPDKWPMPVKAGYLTTPRGAEVLTLMGGFTYDIFRSLGLLAAGCPAEGGYAQAADIDQRT 120
Qy 121 RLTCQAFLDGVAPCGGLTVHNOADLKTDPFLFHPVETGVCKLDNAQTKAIEERLGGPLD 180
Db 121 RLTCQAFLDGVAPCGGLTVHNOADLKTDPFLFHPVETGVCKLDNAQTKAIEERLGGPLD 180
Qy 181 TVSORYAKPFAQMGDVILNFAASPYCKSLQOQKTCDFPAHFAANVNVNKEGTVKVLGGPL 240
Db 181 TVSORYAKPFAQMGDVILNFAASPYCKSLQOQKTCDFPAHFAANVNVNKEGTVKVLGGPL 240
Qy 241 ALSSTIGEITLQNAQAQMPVAVQRLKGAENWVLSLSLHNAQFLNAKTPYIAHKGTP 300
Db 241 ALSSTIGEITLQNAQAQMPVAVQRLKGAENWVLSLSLHNAQFLNAKTPYIAHKGTP 300
Qy 301 LQIDTALTQLDQAQGKLPISAQNRVYFLGGHDTNIAIAGMLGADWQLPPEQDNTPPG 360
Db 301 LQIDTALTQLDQAQGKLPISAQNRVYFLGGHDTNIAIAGMLGADWQLPPEQDNTPPG 360
Qy 361 GGLVFELWQNDHQRVAVKMFYQTMDOIRNAEKLDKKNPAGIISVAVAGCENNGDDK 420
Db 361 GGLVFELWQNDHQRVAVKMFYQTMDOIRNAEKLDKKNPAGIISVAVAGCENNGDDK 420

OM protein - protein search, using sw model
Run on: April 30, 2004, 12:35:05 ; Search time 15.3659 Seconds
(without alignments)
2629.235 Million cell updates/sec
Title: US-10-021-723B-4
Perfect score: 2200
Sequence: 1 MSVLENRVRLSGVLMLSLGL.....NPAGIISVAVAGCENNGDDK 420
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : PIR_78:.*
1: PIR1:.*
2: PIR2:.*
3: PIR3:.*
4: PIR4:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2183	99.2	441	2	AC0201	acid phosphatase (
2	939.5	42.7	432	2	B36733	acid phosphatase (
3	926.5	42.1	434	2	H90770	phosphoanhydride p
4	926.5	42.1	444	2	D85633	hypothetical prote
5	532	24.2	413	2	F90773	periplasmic glucos
6	532	24.2	413	2	B85636	periplasmic glucos
7	530	24.1	413	2	JV0087	glucose-1-phosphat
8	504	22.9	414	2	E87316	periplasmic phosph
9	495	22.5	413	2	AG0632	glucose-1-phosphat
10	462	21.0	417	2	S25627	Glucose-1-phosphat
11	135	5.7	411	2	D88504	protein B0361.7 [i
12	124	5.6	421	2	S14742	acid phosphatase (
13	121.5	5.5	423	1	S06167	acid phosphatase (
14	121	5.5	438	2	S64682	acid phosphatase (
15	115.5	5.2	416	2	T16058	hypothetical prote
16	113	5.1	588	2	AF2542	hypothetical prote
17	108	4.9	26926	1	I38344	titin, cardiac mus
18	107.5	4.9	1940	2	F78393	hypothetical prote
19	106.5	4.8	541	2	S51799	nucleoporin NUP57
20	106	4.8	423	2	A33395	acid phosphatase (
21	105	4.8	468	2	A86233	hypothetical prote
22	103.5	4.7	505	1	PIWLRB	Li protein - cotto
23	102	4.6	537	2	S54770	secreted acid phos
24	102	4.6	888	2	T46726	secreted acid phos
25	101.5	4.6	479	2	F86535	6-phosphogluconate
26	101.5	4.6	479	2	C72088	6-phosphogluconate
27	101.5	4.6	715	2	AC0018	probable membrane
28	100.5	4.6	348	2	AE1751	B. subtilis Cggr h
29	100.5	4.6	840	2	T02164	hypothetical prote

Db 243 EIFLLQQAQMPGCGRITDHSQWNTLLSHNAQFYLLQRTPEVARSATPLLDLIMIA 302
Qy 308 LT---LQIDAQOQKLPISAQRVFLPGCHDNIANIAMGLADWOLPEQDNTPPGGGLV 364
Db 303 LTPHPQQAQVGLTPTS-----VLFIAGHDNLNLAGALELNTLTPGQDNTPPGGGLV 358
Qy 365 FELWONPDNHORYAVKMFYQTMQOLRNAEKLKNNPAGIISVAVAGCE 414
Db 359 FERWERLSDNSQWISVQVSLVFTLQQRDKTFLSL-NTPPGEVKLTLAGCE 407
RESULT 4
D85633
hypothetical protein appA [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: D85633
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: D85633
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-444 <STO>
A:Cross-references: GB:AB005174; NID:G12514245; PIDN:AAG55528.1; GSPDB:GN00145; UWGP:Z.13
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: appA

Query Match 42.1%; Score 926.5; DB 2; Length 444;
Best Local Similarity 47.3%; Pred. No. 5e-67;
Matches 194; Conservative 62; Mismatches 139; Indels 15; Gaps 7;
Qy 13 LVLMGLSLAAITAPVA---ABPS-CYTLERVVLSRHGVRSPTKOTOLMNDVTPDKWPQW 68
Db 15 LIPFLSLIPLTPQSAFQAQSEPELKLKESVIVSRGVRAPTKATQLMQDVTDPANPW 74
Qy 69 PVKAGYLTPRGAEVLTMGFGYGYFRSLGLLA-AGCPAEGGVYAQADIDORTLTGQAF 127
Db 75 PVKGLWLTFRGGLIAYLGHYQRLVADGLLTKKGCPCPCQQAIAIADVDERTKTEGEAF 134
Qy 128 LDGVAPGCGLTVHQAQDLKKTDPLEPHVETGVCKLDNAQTDKAIEERLGGPLDTSQRYA 187
Db 135 AAGLAPDCAITVHQADTSSPDPLFNPLKTVGCQDDNANVTDAIISRAGSIADFTGHRQ 194
Qy 188 KPFAQMGDLNFAASPYCKSLQQQKTCDFAHFAANEVNNKGTKVTLGSLALSSTLG 247
Db 195 TAFRELRVLFNPQSNLCNREKQDESCSLQALPSELKVSAD--NVSLTGAVSLASMLT 252
Qy 248 EIFLLQQAQMPGCGRITDHSQWNTLLSHNAQFYLLQRTPEVARSATPLLDLIMIA 302
Db 253 EIFLLQQAQMPGCGRITDHSQWNTLLSHNAQFYLLQRTPEVARSATPLLDLIMIA 312
Qy 308 LT---LQIDAQOQKLPISAQRVFLPGCHDNIANIAMGLADWOLPEQDNTPPGGGLV 364
Db 313 LTPHPQQAQVGLTPTS-----VLFIAGHDNLNLAGALELNTLTPGQDNTPPGGGLV 368
Qy 365 FELWONPDNHORYAVKMFYQTMQOLRNAEKLKNNPAGIISVAVAGCE 414
Db 369 FERWERLSDNSQWISVQVSLVFTLQQRDKTFLSL-NTPPGEVKLTLAGCE 417

RESULT 5
F90773
periplasmic glucose-1-phosphatase [imported] - Escherichia coli (strain O157:H7, substra
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: F90773
R:Havashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and ger
A:Reference number: A99629; MUID:21156231; PMID:11259796
A:Accession: F90773
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-413 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA834581.1; PID:G13360618; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: ECs1158

Query Match 24.2%; Score 532; DB 2; Length 413;
Best Local Similarity 31.3%; Pred. No. 3.6e-35;
Matches 131; Conservative 83; Mismatches 164; Indels 40; Gaps 11;
Qy 10 LSGVLMLSLAAITAPVAABPSGYTLBRVILSRHGVRSPTKOTOLMNDVTPDKWPQW 68
Db 10 VAGVLLASNAQAQTV-----PEGQLQQLVLMMSRHLNRLAPLANNNGSVLEQSTPNKWPW 64
Qy 69 PVKAGYLTPRGAEVLTMGFGYGYFRSLGLLAAG-CPAEGGVYAQADIDORTLTGQAF 127
Db 65 DVPGGQLTKGGLVLYVMGHYMRLEAQGMVKSGECPDPTVYAYANSQRTVTAQPF 124
Qy 128 LDGVAPGCGLTVHQAQDLKKTDPLEPHVETGVCKLDNAQTDKAIEERLGGPLDTSQRYA 187
Db 125 ITGAFFGCDIPVHHQEKGTMDPTFNPTITDSSAFAFSEQAVAAAMEKELS-----KLQT 178
Qy 188 KPFAQMGDLNFAASPYCKSLQQ-----QKTCDFAHFAANEVNNKGTKVTLGSLALS 243
Db 179 DSVQLLEKIVNKKDSPAECQKQCSLVGDKNT-----FSA-----KYQCEPGVSGFLKVG 228
Qy 244 STLGEIFLLQNAQAMP--EVANQRLKGAENWVLSLHNAQFYLLQRTPEVARSATPLLDLIMIA 301
Db 229 NSLVDAFTLQYVGGPMDQVANGELKSDQKWKSLKNGYQDSLTSPSEVARNVAKPLV 288
Qy 302 QCIDTALTQLDAQOQKLPISAQRVFLPGCHDNIANIAMGLG-ADWOLPEQDNTPPG 360
Db 289 SYIDKALVTDR-----TSAPKITVLVGHDSNIASILLTALDFKPYQLHQNERTPIG 339
Qy 361 GGLVPELWONPDNHORYAVKMFYQTMQOLRNAEKLKNNPAGIISVAVAGCE 418
Db 340 GKIVFORWRDSKANRLMKIEYVYQSAEQLRNADALTQ-APAQRVTLGSLCPDAD 396

RESULT 6
B85636
periplasmic glucose-1-phosphatase [imported] - Escherichia coli (strain O157:H7, subst
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: B85636
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayh
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodac
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: B85636
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-413 <STO>
A:Cross-references: GB:AB005174; NID:G12514273; PIDN:AAG55550.1; GSPDB:GN00145; UWGP:Z.
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: agp

Query Match 24.2%; Score 532; DB 2; Length 413;
Best Local Similarity 31.3%; Pred. No. 3.6e-35;
Matches 131; Conservative 83; Mismatches 164; Indels 40; Gaps 11;
Qy 10 LSGVLMLSLAAITAPVAABPSGYTLBRVILSRHGVRSPTKOTOLMNDVTPDKWPQW 68
Db 10 VAGVLLASNAQAQTV-----PEGQLQQLVLMMSRHLNRLAPLANNNGSVLEQSTPNKWPW 64
Qy 69 PVKAGYLTPRGAEVLTMGFGYGYFRSLGLLAAG-CPAEGGVYAQADIDORTLTGQAF 127

Db 65 DVPGGQLTTKGGVLEVMGYHREWLAEQGWKSGECPDPDVTYVANSIQRVATAQFF 124
Qy 128 LDGVAFCGLTVNQADLKTDPFLPHVETGVCKLQNAQTDKAEERLGGPLDVTVSQRYA 187
Db 125 ITGAFFPGCDIPVHQEKMGMTDFTFNPVITDDSAFSEQAVAAAMEKELS-----KLQLT 178
Qy 188 KPFAQMGDVNLFAASPYCKSLQO---OGKTCDFAHFAANVNNKGGTKVTLGGPLALS 243
Db 179 DSYQLLEKIVNYKDSPACKERKQCSLVDGKNT-----FSA-----KYQEPGVSGPLKV 228
Qy 244 STLGEIFLLQNAQAMP--EVAWQRLKGAENWVLSLHNAQFNLMAKTPVIARHKGTPLL 301
Db 179 DSYQLLEKIVNYKDSPACKERKQCSLVDGKNT-----FSA-----KYQEPGVSGPLKV 228
Qy 244 STLGEIFLLQNAQAMP--EVAWQRLKGAENWVLSLHNAQFNLMAKTPVIARHKGTPLL 301
Db 229 NSLVDATFLQYEGFFMDQVANGIKSDQWKVLSKLKNGYQDSLFTSPEVARNVAKPLV 288
Qy 302 QCIDTALTQLDQAQOKLPIAQNVRVFLGCHDTNIAAGMLG-ADWOLPEQDNTTPG 360
Db 289 SYIDKALVTR-----TSAPKITVLVGHDSNTASLTALDFKPYQLHQNERTPIG 339
Qy 361 GGLVFEWQNDHQRVAVVMFYQTMQDLRNAEKLKNNPAGIISVAVAGCENNGD 418
Db 340 GKIVFQRWDSKANRDLMKIEVYVQSAEQURNADALTQ-APAQRVTLSESGCPIDAD 396

RESULT 7

JV0087

Glucose-1-phosphatase (EC 3.1.3.10) precursor - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C>Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 01-Mar-2002

C:Accession: J00087; H64841

R:Pradel, B.; Marck, C.; Boquet, P.L.

J. Bacteriol. 172, 802-807, 1990

A:Title: Nucleotide sequence and transcriptional analysis of the Escherichia coli agp gene

A:Reference number: JV0087; MUID:90130318; PMID:2153660

A:Accession: JV0087

A:Molecule type: DNA

A:Residues: 1-413 <PRA>

A:CROSS-references: GB:M33807; NID:G145217; PIDN:AAA23426.1; PID:G145218

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: H64841

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-413 <BLAT>

A:CROSS-references: GB:U00096; NID:G1787233; PIDN:AACT4087.1; PID:G1787237;

A:Experimental source: strain K-12, substrain MG1655

C:Genetics:

A:Gene: agp

A:Map position: 23 min

C:Function:

A:Description: essential for growth in a high-phosphate medium containing glucose-1-phosphate

A:Note: optimal at low pH

C:Keywords: homodimer; periplasmic space; phosphohistidine; phosphoprotein; phosphoric acid

F:1-22/Domain: signal sequence #status predicted <SIG>

F:23-413/Product: glucose-1-phosphatase #status predicted <MAT>

F:40/Active site: His (phosphohistidine intermediate) #status predicted

F:311/Active site: His #status predicted

Query Match 24.1%; Score 530; DB 2; Length 413;

Best Local Similarity 31.3%; Pred. No. 5.2e-35;

Matches 131; Conservative 82; Mismatches 165; Indels 40; Gaps 11;

Qy 10 LSGVLVMSGLIAITAPVAAPSGYTLERVVLSRHGVRSP-TKQTLMDNDVTPDKWPOW 68

Db 10 VAGIVLLASNAQAQTV-----PEGYQLQVLLMSRHNLRAFLANNGSVLEQSTPNKWP 64

Qy 69 PVKAGYLTTPGAEVLTLMGFGYDFRSGLGLLAAG-CPAEGGYVQAQADIDQRTLTGQAF 127

Db 65 DVPGGQLTTKGGVLEVMGYHREWLAEQGWKSGECPDPDVTYVANSIQRVATAQFF 124

Qy 128 LDGVAFCGLTVNQADLKTDPFLPHVETGVCKLQNAQTDKAEERLGGPLDVTVSQRYA 187
Db 125 ITGAFFPGCDIPVHQEKMGMTDFTFNPVITDDSAFSEQAVAAAMEKELS-----KLQLT 178
Qy 188 KPFAQMGDVNLFAASPYCKSLQO---OGKTCDFAHFAANVNNKGGTKVTLGGPLALS 243
Db 179 DSYQLLEKIVNYKDSPACKERKQCSLVDGKNT-----FSA-----KYQEPGVSGPLKV 228
Qy 244 STLGEIFLLQNAQAMP--EVAWQRLKGAENWVLSLHNAQFNLMAKTPVIARHKGTPLL 301
Db 229 NSLVDATFLQYEGFFMDQVANGIKSDQWKVLSKLKNGYQDSLFTSPEVARNVAKPLV 288
Qy 302 QCIDTALTQLDQAQOKLPIAQNVRVFLGCHDTNIAAGMLG-ADWOLPEQDNTTPG 360
Db 289 SYIDKALVTR-----TSAPKITVLVGHDSNTASLTALDFKPYQLHQNERTPIG 339
Qy 361 GGLVFEWQNDHQRVAVVMFYQTMQDLRNAEKLKNNPAGIISVAVAGCENNGD 418
Db 340 GKIVFQRWDSKANRDLMKIEVYVQSAEQURNADALTQ-APAQRVTLSESGCPIDAD 396

RESULT 8

E87316

periplasmic phosphoanhydride phosphohydrolase [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001

C:Accession: E87316

R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, B.;

J.; Lamb, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Hatt, D.H.; Kol

n, J.; Emlaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: E87316

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-414 <STO>

A:CROSS-references: GB:AE005673; NID:G13421731; PIDN:AAK22529.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC0542

Query Match 22.9%; Score 504; DB 2; Length 414;

Best Local Similarity 35.0%; Pred. No. 6.7e-33;

Matches 140; Conservative 65; Mismatches 171; Indels 24; Gaps 14;

Qy 20 LAATAPVAA--EPGVTLEVRVILSRHGVSRPTKQTLMDNDVTPDKWPOWPKAGYLTP 77

Db 7 IAVVTLALASGASAELEKLVILSRHGVSRPMSSEERLEASARFPRFVAGHLTA 66

Qy 78 RGAEVLTLMGFGYDFRSGLGLLAAG-CPAEGGYVQAQADIDQRTLTGQAFLDGVAPCG 136

Db 67 RGETLVARMGDIYRRHYAAQGLLXPGDC--ASVYAWANVTQRTIATAKAYRETLAPGCP 123

Qy 137 LTVNQADLKTDPFLPHVETGVCKLQNAQTDKAEERLGGPLDVTVSQRYAKPFAQMGDV 196

Db 124 VTV-NTVGEINIDPFEPVKGIVKADHALARAAGVGGDLTAWASHNQAEQLDAL 182

Qy 197 LNFAASPYCKSLQOQGTCTDFAHFAANVNNKGGTKVTLGGPLALSSTLGEIFLL--QN 254

Db 183 LMQCDKGPAPGKRRV-----FDAKPGFVDGE-ELAGLSGPEAFASGVTESILMAWAD 236

Qy 255 AQAPMEVAVQRLKGAENWVLSLHNAQFNLMAKTPVIARHKGTPFLQCIDTALTQIDA 314

Db 237 GRDFAGLQWKSLL-DEEALTRSFLLHQAEFFDLRLTPPVARTLAGHLADRL--AATLR-DG 292

Qy 315 QGQKLPISAQNRVFLGCHDTNIAAGMLGADQQLP-EQPDNTFPGGGLVFEWQNDP 373

Db 293 AAAIGFVDA--RLVVIAGHDGTLASGLGILLRMENTLPGYQPNQIQPGGALVFERWR--DD 349

Qy 374 HQRYVAVVMFYQTMQDLRNAEKLKNNPAGIISVAVAGC 413

Db 350 GYRVVVRFTQCSLSQLRNMTALDAKTPPLS-APFVQGC 388

RESULT 9

AG0632
Glucose-1-phosphatase precursor (G1Pase), secreted [imported] - Salmonella enterica subsp.
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AG0632
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Comerford, P.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Park, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AG0632
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-413 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD08242.1; PID:G16502289; GSPDB:GN00176
C:Genetics:
A:Gene: STY1153

Query Match 22.5%; Score 495; DB 2; Length 413;
Best Local Similarity 31.8%; Pred. No. 3.6e-32;
Matches 134; Conservative 78; Mismatches 166; Indels 44; Gaps 14;
QY 8 VRLSLGLMLSLGLAATAPVAAPSGYTLERVILSRHGVRSPTKQTLQMDVTPDKWP 66
Db 8 VAVAGAVLLSSAVQAQTT-----PEGYQLQVLMWSRHLAPLANNGNVLAQSTPNWP 62
QY 67 QNPVKAAGYLTTPRGABLVLMGGFYDYPFRSLGLAAG-CPAEGGVYAAQADIDQRTLTGQ 125
Db 63 AWDVPGGLTTKGGVLYVMGYHTRENLVAQGLIPSGECPADPTVYAVANSQRTVATAQ 122
QY 126 AFLDGVAPGGLTTHNQADLKTDPLFHPVETGVCKLDNAQTDKAIEERLGGPLDVTVSOR 185
Db 123 FFIATSAFGCDIPVHQKMGTMDFTFNPVITDSSAATQQAQVAMEK-----ARSQL 175
QY 186 YA-KPFAQMGVDVNFPAASPYCKSLQOQKTCDFAH-----FAANEVNVNKEGTKVTLGSL 240
Db 176 HLDESQYKLEQITHYDQSPSCKEHQ-----CSLIDAKDTFSANTYQ-----EPGVQGL 225
QY 241 ALSSLTGLIFLQNAQAMP--EVAWQRLKGAENWVLSLHNAQFNLMKTPYARHKG 298
Db 226 KVGNSLDAFTLQYIEGFPMDQVAGWGGIHDRQMKVLSKLKNGYDLSFTPTVARNVA 285
QY 299 FLQOQIDTALTQLDAQOQKLPISAQNRVFLGGHDTNANIAGMLG-ADWQLPEQPDNT 357
Db 286 PLVKYIDKVLVAE-----RVSAAP-KVTVLVGHDSNIALTLTALDFKPYQLHDQYERT 336
QY 358 PPGGLVFWLQNPNDHQRVAVVMFYQTMQDLRKAELDLKNPAGIISVAVAGC--EN 415
Db 337 PIGGQLVTFQRWHDGNANRLMKIEYVYQSAQRLRANAEALTLK-SPAQRVTLKGCVPYDA 395
QY 416 NG 417
Db 396 NG 397

RESULT 10

S25627
Glucose-1-phosphatase (EC 3.1.3.10) precursor - Providencia rettgeri
C:Species: Providencia rettgeri
C:Date: 25-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 08-Oct-1999
C:Accession: S25627
R:Riccio, M.B.; Chesurin, A.; Lombardi, G.; Satta, G.
submitted to the EMBL Data Library, September 1992
A:Reference number: S25627
A:Accession: S25627
A:Molecule type: DNA
A:Residues: 1-417 <RIC>
A:Cross-references: EMBL:X68201; NID:G45771; PIDN:CAA48288.1; PID:G45772

A:Experimental source: strain PV7

C:Genetics:
A:Gene: agp
C:Keywords: periplasmic space; phosphohistidine; phosphoprotein; phosphoric monoester
F:1-29/Domain: signal sequence #status predicted <SIG>
F:30-417/Product: glucose-1-phosphatase #status predicted <WAT>
F:43/Active site: His (phosphohistidine intermediate) #status predicted
F:314/Active site: His #status predicted

Query Match 21.0%; Score 462; DB 2; Length 417;
Best Local Similarity 30.6%; Pred. No. 1.7e-29;
Matches 128; Conservative 84; Mismatches 170; Indels 36; Gaps 15;
QY 14 VLMLSGLAATAAPVA-----EPGYTLERVILSRHGVRSPTKQTLQMDVTPDKWP 67
Db 5 VLTLCLSAALFAAPPTMASTNQADMVLDQVILSRHLRTPVNTGILTEVTDKKNPD 64
QY 68 WPKVKAAGYLTTPRGABLVLMGGFYDYPFRSLGLAAG-CPAEG-GVYAAQADIDQRTLTGQ 125
Db 65 WDAKSYLTQGCALLEVYMGHYFREWIDQNKLLADELCPTSNEDIYLVNLSQRTIATAQ 124
QY 126 AFLDGVAPGGLTTHNQADLKTDPLFHPVETGVCKLDNAQTDKAIEERLGGPLDVTVSOR 185
Db 125 FFAAGAPFGCKVNIHQHQPBGKMDPVFNPIITNGSPFKQKALAMDDYLKG-----LSUK 180
QY 186 YAKPFAQMGVDVNFPAASPYCKSLQOQKTCDF--AHFAANEVNVNKEGTKVTLGSLALS 244
Db 181 AG--YEELDTVLNIKDSQCKT----DKLCNLDQSKNSFIIEADKE---PGVSGPLKIAN 231
QY 245 TLGSIFFLQNAQAMP--EVAWQRLKGAENWVLSLHNAQFNLMKTPYARHKGTPILQ 302
Db 232 SAYDAIDLYGEGFPADQVAGLVDTPEKWKLTNKNAYQBTLPFKIIAKNVAHPILN 291
QY 303 QIDTALTQLDAQOQKLPISAQNRVFLGGHDTNANI--AGMLGADWQLPEQPDNTPPGG 361
Db 292 YIDKGF-VSVD-KGETA-----KTFVLVGHDSNIALSAMSAMDPKYQLAQYEHPTIGG 343
QY 362 GLVFELWQNPNDHQRVAVVMFYQTMQDLRKAELDLKNPAGIISVAVAGC--ENNG 417
Db 344 KLVFORWTDQTKKCFMVEYVYQADQLRDNAYLSLETPPKH-VTLELKDQCPVDKNG 400

RESULT 11

D88504
Protein B0361.7 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C:Accession: D88504
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biol
A:Reference number: A75000; MUID:99065613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C_e
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; a
A:Accession: D88504
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-411 <STO>
A:Cross-references: GB:chr_III; PIDN:AAA50626.1; PID:G458955; GSPDB:GN00021; CESP:B036
C:Genetics:
A:Gene: B0361.7
A:Map position: 3
C:Superfamily: mammalian acid phosphatase

Query Match 5.7%; Score 125; DB 2; Length 411;
Best Local Similarity 20.5%; Pred. No. 0.03;
Matches 91; Conservative 58; Mismatches 161; Indels 134; Gaps 20;
QY 13 LVLMLSGLAATA-----PVAAPSGYTLERVILSRHGVRSPTKQTLQMDVTPDKWP 66
Db 4 LLVLLIGASGINAVVYKEVPIQANTD--TLEYVHTVWRHGDRTPAEL-----LPDDIT 55
QY 67 QNPVKAAGYLTTPRGAE-----LVTLMGCFYGDYFRSLGLLAAGCPASGGVYAAQADIDQ 118

[illegible]

RESULT 12

S14742
 acid phosphatase [EC 3.1.3.2] precursor - mouse (fragment)
 CSpecies: Mus musculus (house mouse)
 CDate: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 31-Mar-2000
 CAccession: S14742
 R:Geier, C.; von Figura, K.; Pohlmann, R.
 Biol. Chem. Hoppe-Seyler 372, 301-304, 1991
 ATitle: Molecular cloning of the mouse lysosomal acid phosphatase.
 AReference number: S14742; MUID:91282986; PMID:2059337
 AAccession: S14742
 AStatus: preliminary
 A:Molecule type: mRNA
 AResidues: 1-421 <GEI>
 A:Cross-references: EMBL:X57199; NID:G52870; PIDN:CAA40485.1; PID:G52871
 C:Superfamily: mammalian acid phosphatase
 C:Keywords: lysosome; phosphoric monoester hydrolase

Query Match	5.6%;	Score 124;	DB 2;	Length 421;
Best Local Similarity	22.8%;	Pred. No. 0.037;		
Matches 100;	Conservative 52;	Mismatches 148;	Indels 138;	Gaps 25;
QY	11	SGILVLMISGLA-AITAPVAABPSGYTLERVILSRHGVSPTKOTQLMNDVTPDKWPQW 69		
Db	9	AALQFLLGMCULTVMPPIQAR-----SLRFVTLVLRHGDGRSEVK-TYKPPYQEEKVQ-- 61		
QY	70	VKAGVILTPRGA-----EL-VTLMGGFYG-----DYFRSLGLLAAQCPAEGGYAAQADIQRT 120		
Db	62	-GFGQLTKESMLQHWELGQALRQRYHGFLNYSYHQ-----EYVRSTDFDRT 108		
QY	121	RUTGQAFLDGVAPEGGLTVHQNADLUKTDPLFHVETGVCKVLNQAQTDKAIERLGLPD 180		
Db	109	LMSAEANLAGLFP-----PNEVHFSPNISWQPIP-----VHTVPITEDRL----- 149		
QY	181	TVSQRVAKPFAQMGVDLNAFAASPYCKSLQQQCKT-----CDFAHFAANENVV 227		
Db	150	-----LKPLGFCPRYSQLQNETROTPEYONRSIQNAQPLNVAENETGL 193		
QY	228	NKEGTKVTLSGPLASSTLGEIF-----LLQNAQAMPEVAWQRLKGA--NWVSL 275		
Db	194	-----TNVTL-----EITWVYDTLFCEQTHGLLLPPWASPVQV-QRLSQLKDFSLFL 241		
QY	276	LSLHNAQFNLMAKTPYIARHKGTPTLQIQIDTALTQLDAAQCKLPISAQNRVILFLGHDT 335		
Db	242	FGTH-----EVOVKARLOGCVLLAOLLKNTLIM--ATTSOPF-----KLIVYSAHDT 286		

Db 152 -----LAFPLGCPRYEQLNQETRTQPEYQNESSRNAQFLDMVANE 192
QY 225 VVNKGEKTVLSSGLPLSSSTLGEIFLLQNAQAPPEVAVQELKGAENWVSLSLHNAQFN 284
Db 193 TGL-----TDLTLETWVNYDTL-----FCEQTHGLRLPPW-----ASPQWQRSLRUKDPSFR 241
QY 285 LM-----AKTPYIARHKGTPLLQIDTALTALQDAQCKLPISAQNRVFLGGHDTNIA 341
Db 242 FLFGIYQQAERKARLGGVLLAQIRKNTLM-----ATTSQLP-----KLLVYSAHDTTL--VA 292
QY 342 GMLGADWQLPQPDNTPPGGGLVFLWQNPNDHQRVAVKMFYQT----- 386
Db 293 LQMALDVYNGEQ-----APYASCHIFELYQEDSGN-----FSVEMYFRNESKAPWPLSLPGCP 346
QY 387 -----MDOLRNAE-----KLDLKNPAG-----IISVAVAG 412
Db 347 HRCPLQDFLRLETPVVPKDWQEQCQLASGPADTEVIVALVAGC 389

RESULT 14
S64682
acid phosphatase (EC 3.1.3.2) Acph-1 precursor, lysosomal - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 06-Dec-1996 #sequence_revision 06-Dec-1996 #text_change 17-Mar-2000
C:Accession: S64682; S64681
R:Chung, H.J.; Shafer, C.; MacIntyre, R.
Mol. Gen. Genet. 250, 635-646, 1996
A:Title: Molecular characterization of the lysosomal acid phosphatase from Drosophila melanogaster
A:Reference number: S64681; MUID:96194627; PMID:8676866
A:Accession: S64682
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-438 <CHU>
A:Accession: S64681
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-31, 'Y', 33-438 <CHW>
C:Genetics:
A:Gene: FlyBase:Acph-1
A:Cross-references: FlyBase:FBgn0000032
C:Superfamily: mammalian acid phosphatase
C:Keywords: phosphoric monoester hydrolase
F:1-33/Domain: signal sequence #status predicted <SIG>
F:34-438/Product: acid phosphatase Acph-1 #status predicted <MAT>

Query Match 5.5%; Score 121; DB 2; Length 438;
Best Local Similarity 21.3%; Pred. No. 0.066;
Matches 84; Conservative 60; Mismatches 151; Indels 100; Gaps 19;
QY 36 LERVVILSRHGVSPKQTLQMDVTPDKWPQ---WPVKAGYLTPRGAEVLTLMGGFYCD 92
Db 48 LKPVHVIYRHGDRTP-----VDPYPTDFWGRKFTWGTGMDLTNLGQEHYDLGKWLRLN 101
QY 93 YFRSLGLLAAGCP---AEGGVYQAQ-DIDQRTLTGQAFLDGVAPCGGLTVHNOADLKT 148
Db 102 RYSNL-----LPPIYSNENIYVQSTDVD-RTLMSAQSNLAGLYEPQG-----EDWNT 148
QY 149 DLFPHVETGVCKLDNAQTDKATEERLGGPLDTVSQRYAKPFAQMG-----DVNLFAA 201
Db 149 DINWQPI-----PIHTSPREDPILAAKAPCPAYDELASLES 186
QY 202 SPYCKSLQOQKTCDFAHFAANVNVNKGTVK-TLUGPLALSSLTGEIFLLQNAQAPPE 260
Db 187 SPEFKALTEKHLN-FAILS-----EKGGRPVKFTIDAQYLNNTL-----FIENLYNMTL 235
QY 261 VAM-QRLKGAENWVSLSLHNAQFNMAKTPYIARHKGTPLLQIDTALTALQDAQCKLP 319
Db 236 PKWTKKYYGHE-----ELTYVSNFAFAISSYTRKLARLAKGALLKDI-----FORFKKSSG 287
QY 320 PISAQNRVFLGGHDTNIAAGMLGADWQLPEQPDNTPPGGGLVFLWQNPNDHQRVAV 379
Db 288 SLKPRDSRMWYSADHTTASVLNAL-----KLFLHSP-----YTA 324

QY 380 VKNFYQTMQDLRN-AEKLDLKNPAGIISVAVAGC 413
Db 325 CIMMELRVDBTNTPLVSIYFYKNTTAEPLEPLDIPGC 359
RESULT 15
T16058
hypothetical protein F13D11.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 23-Mar-2001
C:Accession: T16058
R:Fulton, L.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid F13D11.
A:Reference number: S69020
A:Accession: T16058
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-416 <FUL>
A:Cross-references: EMBL:U40939; NID:G1072175; PID:G1072178; PIDN:AAA61702.1; CESP:F13
C:Genetics:
A:Gene: CESP:F13D11.1
A:Introns: 23/3; 59/3; 88/3; 138/3; 202/3; 244/1; 270/3; 312/3; 349/3; 377/1
C:Superfamily: mammalian acid phosphatase

Query Match 5.2%; Score 115.5; DB 2; Length 416;
Best Local Similarity 21.8%; Pred. No. 0.18;
Matches 90; Conservative 60; Mismatches 118; Indels 149; Gaps 25;
QY 13 LVLMLSGLAITAPVAAPSPGYTLERWILSRHGVSPKQTLQ-----MNDVTPDK--WP 66
Db 3 LVLLLPFL-----FPVAFG-----KLKFWQIWFHRTGPHYLYFPDGLNNVDYQQIAMP 54
QY 67 QWPVKAGYLTPRGA-----ELVTLMGFGYGDYFSLGLAAGCPAEGGVYAAADI----- 116
Db 55 -----GELTKRGILBEFQLGQELRKIYGEHF-----GDTYQPRDFHYVTGK 95
QY 117 DQRTLTGQAFLDGVAPCGGLTVHNOADLKTDPDFHVPETGVCKLDNAQ-TDKATEERL 175
Db 96 DNRTSASQAQAMPAGFLP-----PNEDQTNWYELKWPV-----AQLTDESIDWVS 140
QY 176 GGFLDTV-----SQRYAKPFAQMGD-----VLNFAASPYCKSLQOQKTCDF 218
Db 141 LGAIDNCPVYGEAQRKSEYAEVMDQEKYDABELLQVRNHADEPIVEAVK-----YN 193
QY 219 HFAANVNVNKGTVKTLGSLALSSLTGEIFLLQNAQAPPEVAVQELKGAENWVSLLS- 277
Db 194 H-----VIDSLKVRVILQD-DRLFPYEWAR--GYENRIINMSF 228
QY 278 LHNQAQNLMAKTPYIARHKGTPLLQIDTALTALQDAQCKLPISAQNRVFLGGHDTN 336
Db 229 LIHDAV--VKVQNDVSGDYHNLVMSYFETHL-----OK-----NSTKGVFISGHD 274
QY 337 IANTAGMLGADWQ---LPEQPDNTPPGGGLVFLWQNPNDHQRVAVKMFYQTMQDL 390
Db 275 LVTI-----WESLRIDGHPEIDP-----NYGAHIALEM-HEPVGQL 309

Search completed: April 30, 2004, 12:42:18
Job time : 18.3659 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 30, 2004, 12:31:10 ; Search time 10.2439 Seconds
(without alignments)
2134.876 Million cell updates/sec

Title: US-10-021-723B-4
Perfect score: 2200
Sequence: 1 MSVLENRVRLSGLVLMISGL.....NPAGIISVAVACENNGDDX 420

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
1	939.5	42.7	432	1	PPA_ECOLI	P07102	escherichia
2	530	24.1	413	1	AGP_ECOLI	P19926	escherichia
3	499	22.7	413	1	AGP_SALTY	O39921	salmonella
4	462	21.0	417	1	AGP_PRORE	Q52309	providencia
5	125	5.7	422	1	PPAY_CABEL	Q10944	caenorhabdi
6	124	5.6	423	1	PPAL_MOUSE	P24638	mus musculus
7	121.5	5.5	423	1	PPAL_HUMAN	P11117	homo sapien
8	106.5	4.8	541	1	NU57_YEAST	P48837	saccharomyc
9	106	4.8	423	1	PPAL_RAT	P20611	rattus norv
10	103.5	4.7	505	1	VLI_CRPVK	P03102	cottontail
11	101.5	4.6	479	1	6PGD_CHLEP	Q92813	chlamydia p
12	101.5	4.6	715	1	IGAA_YERPE	P58722	yersinia pe
13	99.5	4.5	4367	1	DVHC_NEUCR	P45443	neurospora
14	98.5	4.5	411	1	VIB2_HCMVA	P19893	human cytom
15	98.5	4.5	586	1	CYDD_HAFIN	P45082	haemophilus
16	98.5	4.5	609	1	VGI6_BPP22	Q01146	bacterioph
17	97.5	4.4	1275	1	RFSC_MYXIA	Q50864	myxococcus
18	95	4.3	743	1	FURL_RHIME	Q92977	rhizobium m
19	95	4.3	4485	1	DYHG_CHLRE	Q39575	chlamydomon
20	94.5	4.3	1049	1	EF3_CANAL	P25997	candida alb
21	93.5	4.2	1584	1	KYK1_DICDI	P18160	dictyosteli
22	92.5	4.2	410	1	VIB2_HCMVT	P06435	human cytom
23	92	4.2	513	1	F752_SOLME	P31120	solanum mel
24	91.5	4.2	666	1	FPCK_WAIZE	Q95120	zea mays (m
25	91.5	4.2	720	1	FPTA_PSEAE	P42512	pseudomonas
26	91.5	4.2	772	1	TLK2_HUMAN	Q86ue8	homo sapien
27	91	4.1	390	1	MALY_ECOLI	P23256	escherichia
28	91	4.1	487	1	MURE_PSES	Q87wy0	pseudomonas
29	91	4.1	506	1	C751_PETHY	P48418	perutunia hyb
30	91	4.1	745	1	PO21_EIG	Q29076	sus scrofa
31	90.5	4.1	508	1	FSBB_EUGGR	P14813	euglena gra
32	90.5	4.1	1063	1	HGPI_HABIN	P44795	haemophilus
33	90	4.1	381	1	PPAP_RAT	P20646	rattus norv

Q8ywf0 anabaena sp
Q8euy3 mycoplasma
Q08i36 cercocobus
P55540 rhizobium s
P52732 homo sapien
Q03132 saccharopol
P48419 perutunia hyb
P09489 serratia ma
Q55669 synecocyst
P00545 feline sarc
P79184 macaca fusc
Q08340 macaca neme

ALIGNMENTS

RESULT 1
PPA_ECOLI
ID_PPA_ECOLI STANDARD; PRT; 432 AA.
AC P07102;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Periplasmic appa protein precursor [Includes: Phosphoanhydride
DE phosphohydrolase (EC 3.1.3.2) (pH 2.5 acid phosphatase) (AP); 4-
DE phytase (EC 3.1.3.26)].
GN APPA OR B0980
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-29.
RC STRAIN=K12;
RX MEDLINE=90368616; PubMed=2168385;
RA Dassa J., Marck C., Soquet F.L.;
RT "The complete nucleotide sequence of the Escherichia coli gene appa
RT reveals significant homology between pH 2.5 acid phosphatase and
RT glucose-1-phosphatase."
RL J. Bacteriol. 172:5497-5500(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Kikuchi K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map."
RL DNA Res. 3:137-155(1996).
RN [4]
RP SEQUENCE OF 1-112 FROM N.A.
RX MEDLINE=87271766; PubMed=3038201;
RA Touati E., Danchin A.;
RT "The structure of the promoter and amino terminal region of the pH
RT 2.5 acid phosphatase structural gene (appa) of E. coli: a negative
RT control of transcription mediated by cyclic AMP."
RL Biochimie 69:215-221(1987).
RN [5]

SEQUENCE OF 1-17 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=92049231; PubMed=1658595;
RA Dassa J., Fsihi H., Marck C., Dion M., Kieffer-Bontemps M.,
RA Boquet P.B.;
RT "A new oxygen-regulated operon in *Escherichia coli* comprises the
RT genes for a putative third cytochrome oxidase and for pH 2.5 acid
RT phosphatase (appa).";
RL Mol. Gen. Genet. 229:341-352(1991).
RN [6]
RP CHARACTERIZATION, AND SEQUENCE OF 23-34.
RA Greiner R., Jany K.D.;
RT "Characterization of a phytase from *Escherichia coli*.";
RL Biol. Chem. Hoppe-Seyler 372:664-665(1991).
RN [7]
RP CHARACTERIZATION, AND SEQUENCE OF 23-35.
RX MEDLINE=93256556; PubMed=8387749;
RA Greiner R., Konietzny U., Jany K.D.;
RT "Purification and characterization of two phytases from *Escherichia coli*.";
RL Arch. Biochem. Biophys. 303:107-113(1993).
RN [8]
RP CHARACTERIZATION.
RX MEDLINE=20161462; PubMed=10696472;
RA Golovan S., Wang G., Zhang J., Forsberg C.W.;
RT "Characterization and overproduction of the *Escherichia coli* appa
RT encoded bifunctional enzyme that exhibits both phytase and acid
RT phosphatase activities.";
RL Can. J. Microbiol. 46:59-71(2000).
RN [9]
RP MUTAGENESIS.
RX MEDLINE=93054596; PubMed=1429631;
RA Ostanin K., Harms E.H., Stevis P.E., Kuciel R., Zhou M.-M.,
RA van Etten R.L.;
RT "Overexpression, site-directed mutagenesis, and mechanism of
RT *Escherichia coli* acid phosphatase.";
RL J. Biol. Chem. 267:22830-22836(1992).
RN [10]
RP X-RAY CRYSTALLOGRAPHY (2.05 ANGSTROMS).
RX MEDLINE=20122624; PubMed=10655611;
RA Lim D., Golovan S., Forsberg C.W., Jia Z.;
RT "Crystal structures of *Escherichia coli* phytase and its complex with
RT phytate.";
RL Nat. Struct. Biol. 7:108-113(2000).
CC -!- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O = an
CC alcohol + phosphate.
CC -!- CATALYTIC ACTIVITY: Myo-inositol hexakisphosphate + H(2)O = 1D-
CC myo-inositol 1,2,3,4,5-pentakisphosphate + phosphate.
CC -!- SUBUNIT: Monomer.
CC -!- SUBCELLULAR LOCATION: Periplasmic.
CC -!- INDUCTION: IN ADDITION TO CAMP-MEDIATED CONTROL, THIS ENZYME IS
CC INDUCED WHEN BACTERIAL CULTURES REACH STATIONARY PHASE; ITS
CC SYNTHESIS IS TRIGGERED BY PHOSPHATE STARVATION OR A SHIFT FROM
CC AEROBIC TO ANAEROBIC CONDITIONS.
CC -!- SIMILARITY: Belongs to the histidine acid phosphatase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M58708; AAB20286.1; -;
DR EMBL; AE000200; AAC74065.1; -;
DR EMBL; D90735; BAA35745.1; -;
DR EMBL; X05471; CAA29031.1; -;
DR EMBL; M58708; -; NOT ANNOTATED_CDS.
DR EMBL; S63811; AAB20286.1; -;
DR F01; B36733; B36733.
DR PDB; 1DKL; 03-AUG-00.
DR PDB; 1DKM; 02-AUG-00.

DR PDB; 1DKN; 03-AUG-00.
DR PDB; 1DKO; 03-AUG-00.
DR PDB; 1DKP; 03-AUG-00.
DR PDB; 1DKQ; 03-AUG-00.
DR PDB; 1DKR; 03-AUG-00.
DR PDB; 1DKS; 03-AUG-00.
DR PDB; 1DKT; 03-AUG-00.
DR PDB; 1DKU; 03-AUG-00.
DR PDB; 1DKV; 03-AUG-00.
DR PDB; 1DKW; 03-AUG-00.
DR PDB; 1DKX; 03-AUG-00.
DR PDB; 1DKY; 03-AUG-00.
DR PDB; 1DKZ; 03-AUG-00.
DR PDB; 1DK1; 03-AUG-00.
DR PDB; 1DK2; 03-AUG-00.
DR PDB; 1DK3; 03-AUG-00.
DR PDB; 1DK4; 03-AUG-00.
DR PDB; 1DK5; 03-AUG-00.
DR PDB; 1DK6; 03-AUG-00.
DR PDB; 1DK7; 03-AUG-00.
DR PDB; 1DK8; 03-AUG-00.
DR PDB; 1DK9; 03-AUG-00.
DR PDB; 1DKA; 03-AUG-00.
DR PDB; 1DKB; 03-AUG-00.
DR PDB; 1DKC; 03-AUG-00.
DR PDB; 1DKD; 03-AUG-00.
DR PDB; 1DKE; 03-AUG-00.
DR PDB; 1DKF; 03-AUG-00.
DR PDB; 1DKG; 03-AUG-00.
DR PDB; 1DKH; 03-AUG-00.
DR PDB; 1DKI; 03-AUG-00.
DR PDB; 1DKJ; 03-AUG-00.
DR PDB; 1DKK; 03-AUG-00.
DR PDB; 1DKL; 03-AUG-00.
DR PDB; 1DKM; 03-AUG-00.
DR PDB; 1DKN; 03-AUG-00.
DR PDB; 1DKO; 03-AUG-00.
DR PDB; 1DKP; 03-AUG-00.
DR PDB; 1DKQ; 03-AUG-00.
DR PDB; 1DKR; 03-AUG-00.
DR PDB; 1DKS; 03-AUG-00.
DR PDB; 1DKT; 03-AUG-00.
DR PDB; 1DKU; 03-AUG-00.
DR PDB; 1DKV; 03-AUG-00.
DR PDB; 1DKW; 03-AUG-00.
DR PDB; 1DKX; 03-AUG-00.
DR PDB; 1DKY; 03-AUG-00.
DR PDB; 1DKZ; 03-AUG-00.
DR PDB; 1DK1; 03-AUG-00.
DR PDB; 1DK2; 03-AUG-00.
DR PDB; 1DK3; 03-AUG-00.
DR PDB; 1DK4; 03-AUG-00.
DR PDB; 1DK5; 03-AUG-00.
DR PDB; 1DK6; 03-AUG-00.
DR PDB; 1DK7; 03-AUG-00.
DR PDB; 1DK8; 03-AUG-00.
DR PDB; 1DK9; 03-AUG-00.
DR PDB; 1DKA; 03-AUG-00.
DR PDB; 1DKB; 03-AUG-00.
DR PDB; 1DKC; 03-AUG-00.
DR PDB; 1DKD; 03-AUG-00.
DR PDB; 1DKE; 03-AUG-00.
DR PDB; 1DKF; 03-AUG-00.
DR PDB; 1DKG; 03-AUG-00.
DR PDB; 1DKH; 03-AUG-00.
DR PDB; 1DKI; 03-AUG-00.
DR PDB; 1DKJ; 03-AUG-00.
DR PDB; 1DKK; 03-AUG-00.
DR PDB; 1DKL; 03-AUG-00.
DR PDB; 1DKM; 03-AUG-00.
DR PDB; 1DKN; 03-AUG-00.
DR PDB; 1DKO; 03-AUG-00.
DR PDB; 1DKP; 03-AUG-00.
DR PDB; 1DKQ; 03-AUG-00.
DR PDB; 1DKR; 03-AUG-00.
DR PDB; 1DKS; 03-AUG-00.
DR PDB; 1DKT; 03-AUG-00.
DR PDB; 1DKU; 03-AUG-00.
DR PDB; 1DKV; 03-AUG-00.
DR PDB; 1DKW; 03-AUG-00.
DR PDB; 1DKX; 03-AUG-00.
DR PDB; 1DKY; 03-AUG-00.
DR PDB; 1DKZ; 03-AUG-00.
DR PDB; 1DK1; 03-AUG-00.
DR PDB; 1DK2; 03-AUG-00.
DR PDB; 1DK3; 03-AUG-00.
DR PDB; 1DK4; 03-AUG-00.
DR PDB; 1DK5; 03-AUG-00.
DR PDB; 1DK6; 03-AUG-00.
DR PDB; 1DK7; 03-AUG-00.
DR PDB; 1DK8; 03-AUG-00.
DR PDB; 1DK9; 03-AUG-00.
DR PDB; 1DKA; 03-AUG-00.
DR PDB; 1DKB; 03-AUG-00.
DR PDB; 1DKC; 03-AUG-00.
DR PDB; 1DKD; 03-AUG-00.
DR PDB; 1DKE; 03-AUG-00.
DR PDB; 1DKF; 03-AUG-00.
DR PDB; 1DKG; 03-AUG-00.
DR PDB; 1DKH; 03-AUG-00.
DR PDB; 1DKI; 03-AUG-00.
DR PDB; 1DKJ; 03-AUG-00.
DR PDB; 1DKK; 03-AUG-00.
DR PDB; 1DKL; 03-AUG-00.
DR PDB; 1DKM; 03-AUG-00.
DR PDB; 1DKN; 03-AUG-00.
DR PDB; 1DKO; 03-AUG-00.
DR PDB; 1DKP; 03-AUG-00.
DR PDB; 1DKQ; 03-AUG-00.
DR PDB; 1DKR; 03-AUG-00.
DR PDB; 1DKS; 03-AUG-00.
DR PDB; 1DKT; 03-AUG-00.
DR PDB; 1DKU; 03-AUG-00.
DR PDB; 1DKV; 03-AUG-00.
DR PDB; 1DKW; 03-AUG-00.
DR PDB; 1DKX; 03-AUG-00.
DR PDB; 1DKY; 03-AUG-00.
DR PDB; 1DKZ; 03-AUG-00.
DR PDB; 1DK1; 03-AUG-00.
DR PDB; 1DK2; 03-AUG-00.
DR PDB; 1DK3; 03-AUG-00.
DR PDB; 1DK4; 03-AUG-00.
DR PDB; 1DK5; 03-AUG-00.
DR PDB; 1DK6; 03-AUG-00.
DR PDB; 1DK7; 03-AUG-00.
DR PDB; 1DK8; 03-AUG-00.
DR PDB; 1DK9; 03-AUG-00.
DR PDB; 1DKA; 03-AUG-00.
DR PDB; 1DKB; 03-AUG-00.
DR PDB; 1DKC; 03-AUG-00.
DR PDB; 1DKD; 03-AUG-00.
DR PDB; 1DKE; 03-AUG-00.
DR PDB; 1DKF; 03-AUG-00.
DR PDB; 1DKG; 03-AUG-00.
DR PDB; 1DKH; 03-AUG-00.
DR PDB; 1DKI; 03-AUG-00.
DR PDB; 1DKJ; 03-AUG-00.
DR PDB; 1DKK; 03-AUG-00.
DR PDB; 1DKL; 03-AUG-00.
DR PDB; 1DKM; 03-AUG-00.
DR PDB; 1DKN; 03-AUG-00.
DR PDB; 1DKO; 03-AUG-00.
DR PDB; 1DKP; 03-AUG-00.
DR PDB; 1DKQ; 03-AUG-00.
DR PDB; 1DKR; 03-AUG-00.
DR PDB; 1DKS; 03-AUG-00.
DR PDB; 1DKT; 03-AUG-00.
DR PDB; 1DKU; 03-AUG-00.
DR PDB; 1DKV; 03-AUG-00.
DR PDB; 1DKW; 03-AUG-00.
DR PDB; 1DKX; 03-AUG-00.
DR PDB; 1DKY; 03-AUG-00.
DR PDB; 1DKZ; 03-AUG-00.
DR PDB; 1DK1; 03-AUG-00.
DR PDB; 1DK2; 03-AUG-00.
DR PDB; 1DK3; 03-AUG-00.
DR PDB; 1DK4; 03-AUG-00.
DR PDB; 1DK5; 03-AUG-00.
DR PDB; 1DK6; 03-AUG-00.
DR PDB; 1DK7; 03-AUG-00.
DR PDB; 1DK8; 03-AUG-00.
DR PDB; 1DK9; 03-AUG-00.
DR PDB; 1DKA; 03-AUG-00.
DR PDB; 1DKB; 03-AUG-00.
DR PDB; 1DKC; 03-AUG-00.
DR PDB; 1DKD; 03-AUG-00.
DR PDB; 1DKE; 03-AUG-00.
DR PDB; 1DKF; 03-AUG-00.
DR PDB; 1DKG; 03-AUG-00.
DR PDB; 1DKH; 03-AUG-00.
DR PDB; 1DKI; 03-AUG-00.
DR PDB; 1DKJ; 03-AUG-00.
DR PDB; 1DKK; 03-AUG-00.
DR PDB; 1DKL; 03-AUG-00.
DR PDB; 1DKM; 03-AUG-00.
DR PDB; 1DKN; 03-AUG-00.
DR PDB; 1DKO; 03-AUG-00.
DR PDB; 1DKP; 03-AUG-00.
DR PDB; 1DKQ; 03-AUG-00.
DR PDB; 1DKR; 03-AUG-00.
DR PDB; 1DKS; 03-AUG-00.
DR PDB; 1DKT; 03-AUG-00.
DR PDB; 1DKU; 03-AUG-00.
DR PDB; 1DKV; 03-AUG-00.
DR PDB; 1DKW; 03-AUG-00.
DR PDB; 1DKX; 03-AUG-00.
DR PDB; 1DKY; 03-AUG-00.
DR PDB; 1DKZ; 03-AUG-00.
DR PDB; 1DK1; 03-AUG-00.
DR PDB; 1DK2; 03-AUG-00.
DR PDB; 1DK3; 03-AUG-00.
DR PDB; 1DK4; 03-AUG-00.
DR PDB; 1DK5; 03-AUG-00.
DR PDB; 1DK6; 03-AUG-00.
DR PDB; 1DK7; 03-AUG-00.
DR PDB; 1DK8; 03-AUG-00.
DR PDB; 1DK9; 03-AUG-00.
DR PDB; 1DKA; 03-AUG-00.
DR PDB; 1DKB; 03-AUG-00.
DR PDB; 1DKC; 03-AUG-00.
DR PDB; 1DKD; 03-AUG-00.
DR PDB; 1DKE; 03-AUG-00.
DR PDB; 1DKF; 03-AUG-00.
DR PDB; 1DKG; 03-AUG-00.
DR PDB; 1DKH; 03-AUG-00.
DR PDB; 1DKI; 03-AUG-00.
DR PDB; 1DKJ; 03-AUG-00.
DR PDB; 1DKK; 03-AUG-00.
DR PDB; 1DKL; 03-AUG-00.
DR PDB; 1DKM; 03-AUG-00.
DR PDB; 1DKN; 03-AUG-00.
DR PDB; 1DKO; 03-AUG-00.
DR PDB; 1DKP; 03-AUG-00.
DR PDB; 1DKQ; 03-AUG-00.
DR PDB; 1DKR; 03-AUG-00.
DR PDB; 1DKS; 03-AUG-00.
DR PDB; 1DKT; 03-AUG-00.
DR PDB; 1DKU; 03-AUG-00.
DR PDB; 1DKV; 03-AUG-00.
DR PDB; 1DKW; 03-AUG-00.
DR PDB; 1DKX; 03-AUG-00.
DR PDB; 1DKY; 03-AUG-00.
DR PDB; 1DKZ; 03-AUG-00.
DR PDB; 1DK1; 03-AUG-00.
DR PDB; 1DK2; 03-AUG-00.
DR PDB; 1DK3; 03-AUG-00.
DR PDB; 1DK4; 03-AUG-00.
DR PDB; 1DK5; 03-AUG-00.
DR PDB; 1DK6; 03-AUG-00.
DR PDB; 1DK7; 03-AUG-00.
DR PDB; 1DK8; 03-AUG-00.
DR PDB; 1DK9; 03-AUG-00.
DR PDB; 1DKA; 03-AUG-00.
DR PDB; 1DKB; 03-AUG-00.
DR PDB; 1DKC; 03-AUG-00.
DR PDB; 1DKD; 03-AUG-00.
DR PDB; 1DKE; 03-AUG-00.
DR PDB; 1DKF; 03-AUG-00.
DR PDB; 1DKG; 03-AUG-00.
DR PDB; 1DKH; 03-AUG-00.
DR PDB; 1DKI; 03-AUG-00.
DR PDB; 1DKJ; 03-AUG-00.
DR PDB; 1DKK; 03-AUG-00.
DR PDB; 1DKL; 03-AUG-00.
DR PDB; 1DKM; 03-AUG-00.
DR PDB; 1DKN; 03-AUG-00.
DR PDB; 1DKO; 03-AUG-00.
DR PDB; 1DKP; 03-AUG-00.
DR PDB; 1DKQ; 03-AUG-00.
DR PDB; 1DKR; 03-AUG-00.
DR PDB; 1DKS; 03-AUG-00.
DR PDB; 1DKT; 03-AUG-00.
DR PDB; 1DKU; 03-AUG-00.
DR PDB; 1DKV; 03-AUG-00.
DR PDB; 1DKW; 03-AUG-00.
DR PDB; 1DKX; 03-AUG-00.
DR PDB; 1DKY; 03-AUG-00.
DR PDB; 1DKZ; 03-AUG-00.
DR PDB; 1DK1; 03-AUG-00.
DR PDB; 1DK2; 03-AUG-00.
DR PDB; 1DK3; 03-AUG-00.
DR PDB; 1DK4; 03-AUG-00.
DR PDB; 1DK5; 03-AUG-00.
DR PDB; 1DK6; 03-AUG-00.
DR PDB; 1DK7; 03-AUG-00.
DR PDB; 1DK8; 03-AUG-00.
DR PDB; 1DK9; 03-AUG-00.
DR PDB; 1DKA; 03-AUG-00.
DR PDB; 1DKB; 03-AUG-00.
DR PDB; 1DKC; 03-AUG-00.
DR PDB; 1DKD; 03-AUG-00.
DR PDB; 1DKE; 03-AUG-00.
DR PDB; 1DKF; 03-AUG-00.
DR PDB; 1DKG; 03-AUG-00.
DR PDB; 1DKH; 03-AUG-00.
DR PDB; 1DKI; 03-AUG-00.
DR PDB; 1DKJ; 03-AUG-00.
DR PDB; 1DKK; 03-AUG-00.
DR PDB; 1DKL; 03-AUG-00.
DR PDB; 1DKM; 03-AUG-00.
DR PDB; 1DKN; 03-AUG-00.
DR PDB; 1DKO; 03-AUG-00.
DR PDB; 1DKP; 03-AUG-00.
DR PDB; 1DKQ; 03-AUG-00.
DR PDB; 1DKR; 03-AUG-00.
DR PDB; 1DKS; 03-AUG-00.
DR PDB; 1DKT; 03-AUG-00.
DR PDB; 1DKU; 03-AUG-00.
DR PDB; 1DKV; 03-AUG-00.
DR PDB; 1DKW; 03-AUG-00.
DR PDB; 1DKX; 03-AUG-00.
DR PDB; 1DKY; 03-AUG-00.
DR PDB; 1DKZ; 03-AUG-00.
DR PDB; 1DK1; 03-AUG-00.
DR PDB; 1DK2; 03-AUG-00.
DR PDB; 1DK3; 03-AUG-00.
DR PDB; 1DK4; 03-AUG-00.
DR PDB; 1DK5; 03-AUG-00.
DR PDB; 1DK6; 03-AUG-00.
DR PDB; 1DK7; 03-AUG-00.
DR PDB; 1DK8; 03-AUG-00.
DR PDB; 1DK9; 03-AUG-00.
DR PDB; 1DKA; 03-AUG-00.
DR PDB; 1DKB; 03-AUG-00.
DR PDB; 1DKC; 03-AUG-00.
DR PDB; 1DKD; 03-AUG-00.
DR PDB; 1DKE; 03-AUG-00.
DR PDB; 1DKF; 03-AUG-00.
DR PDB; 1DKG; 03-AUG-00.
DR PDB; 1DKH; 03-AUG-00.
DR PDB; 1DKI; 03-AUG-00.
DR PDB; 1DKJ; 03-AUG-00.
DR PDB; 1DKK; 03-AUG-00.
DR PDB; 1DKL; 03-AUG-00.
DR PDB; 1DKM; 03-AUG-00.
DR PDB; 1DKN; 03-AUG-00.
DR PDB; 1DKO; 03-AUG-00.
DR PDB; 1DKP; 03-AUG-00.
DR PDB; 1DKQ; 03-AUG-00.
DR PDB; 1DKR; 03-AUG-00.
DR PDB; 1DKS; 03-AUG-00.
DR PDB; 1DKT; 03-AUG-00.
DR PDB; 1DKU; 03-AUG-00.
DR PDB; 1DKV; 03-AUG-00.
DR PDB; 1DKW; 03-AUG-00.
DR PDB; 1DKX; 03-AUG-00.
DR PDB; 1DKY; 03-AUG-00.
DR PDB; 1DKZ; 03-AUG-00.
DR PDB; 1DK1; 03-AUG-00.
DR PDB; 1DK2; 03-AUG-00.
DR PDB; 1DK3; 03-AUG-00.
DR PDB; 1DK4; 03-AUG-00.
DR PDB; 1DK5; 03-AUG-00.
DR PDB; 1DK6; 03-AUG-00.
DR PDB; 1DK7; 03-AUG-00.
DR PDB; 1DK8; 03-AUG-00.
DR PDB; 1DK9; 03-AUG-00.
DR PDB; 1DKA; 03-AUG-00.
DR PDB; 1DKB; 03-AUG-00.
DR PDB; 1DKC; 03-AUG-00.
DR PDB; 1DKD; 03-AUG-00.
DR PDB; 1DKE; 03-AUG-00.
DR PDB; 1DKF; 03-AUG-00.
DR PDB; 1DKG; 03-AUG-00.
DR PDB; 1DKH; 03-AUG-00.
DR PDB; 1DKI; 03-AUG-00.
DR PDB; 1DKJ; 03-AUG-00.
DR PDB; 1DKK; 03-AUG-00.
DR PDB; 1DKL; 03-AUG-00.
DR PDB; 1DKM; 03-AUG-00.
DR PDB; 1DKN; 03-AUG-00.
DR PDB; 1DKO; 03-AUG-00.
DR PDB; 1DKP; 03-AUG-00.
DR PDB; 1DKQ; 03-AUG-00.
DR PDB; 1DKR; 03-AUG-00.
DR PDB; 1DKS; 03-AUG-00.
DR PDB; 1DKT; 03-AUG-00.
DR PDB; 1DKU; 03-AUG-00.
DR PDB; 1DKV; 03-AUG-00.
DR PDB; 1DKW; 03-AUG-00.
DR PDB; 1DKX; 03-AUG-00.
DR PDB; 1DKY; 03-AUG-00.
DR PDB; 1DKZ; 03-AUG-00.
DR PDB; 1DK1; 03-AUG-00.
DR PDB; 1DK2; 03-AUG-00.
DR PDB; 1DK3; 03-AUG-00.
DR PDB; 1DK4; 03-AUG-00.
DR PDB; 1DK5; 03-AUG-00.
DR PDB; 1DK6; 03-AUG-00.
DR PDB; 1DK7; 03-AUG-00.
DR PDB; 1DK8; 03-AUG-00.
DR PDB; 1DK9; 03-AUG-00.
DR PDB; 1DKA; 03-AUG-00.
DR PDB; 1DKB; 03-AUG-00.
DR PDB; 1DKC; 03-AUG-00.
DR PDB; 1DKD; 03-AUG-00.
DR PDB; 1DKE; 03-AUG-00.
DR PDB; 1DKF; 03-AUG-00.
DR PDB; 1DKG; 03-AUG-00.
DR PDB; 1DKH; 03-AUG-00.
DR PDB; 1DKI; 03-AUG-00.
DR PDB; 1DKJ; 03-AUG-00.
DR PDB; 1DKK; 03-AUG-00.
DR PDB; 1DKL; 03-AUG-00.
DR PDB; 1DKM; 03-AUG-00.
DR PDB; 1DKN; 03-AUG-00.
DR PDB; 1DKO; 03-AUG-00.
DR PDB; 1DKP; 03-AUG-00.
DR PDB; 1DKQ; 03-AUG-00.
DR PDB; 1DKR; 03-AUG-00.
DR PDB; 1DKS; 03-AUG-00.
DR PDB; 1DKT; 03-AUG-00.
DR PDB; 1DKU; 03-AUG-00.
DR PDB; 1DKV; 03-AUG-00.
DR PDB; 1DKW; 03-AUG-00.
DR PDB; 1DKX; 03-AUG-00.
DR PDB; 1DKY; 03-AUG-00.
DR PDB; 1DKZ; 03-AUG-00.
DR PDB; 1DK1; 03-AUG-00.
DR PDB; 1DK2; 03-AUG-00.
DR PDB; 1DK3; 03-AUG-00.
DR PDB; 1DK4; 03-AUG-00.
DR PDB; 1DK5; 03-AUG-00.
DR PDB; 1DK6; 03-AUG-00.
DR PDB; 1DK7; 03-AUG-00.
DR PDB; 1DK8; 03-AUG-00.
DR PDB; 1DK9; 03-AUG-00.
DR PDB; 1DKA; 03-AUG-00.
DR PDB; 1DKB; 03-AUG-00.
DR PDB; 1DKC; 03-AUG-00.
DR PDB; 1DKD; 03-AUG-00.
DR PDB; 1DKE; 03-AUG-00.
DR PDB; 1DKF; 03-AUG-00.
DR PDB; 1DKG; 03-AUG-00.
DR PDB; 1DKH; 03-AUG-00.
DR PDB; 1DKI; 03-AUG-00.
DR PDB; 1DKJ; 03-AUG-00.
DR PDB; 1DKK; 03-AUG-00.
DR PDB; 1DKL; 03-AUG-00.
DR PDB; 1DKM; 03-AUG-00.
DR PDB; 1DKN; 03-AUG-00.
DR PDB; 1DKO; 03-AUG-00.
DR PDB; 1DKP; 03-AUG-00.
DR PDB; 1DKQ; 03-AUG-00.
DR PDB; 1DKR; 03-AUG-00.
DR PDB; 1DKS; 03-AUG-00.
DR PDB; 1DKT; 03-AUG-00.
DR PDB; 1DKU; 03-AUG-00.
DR PDB; 1DKV; 03-AUG-00.
DR PDB; 1DKW; 03-AUG-00.
DR PDB; 1DKX; 03-AUG-00.
DR PDB; 1DKY; 03-AUG-00.
DR PDB; 1DKZ; 03-AUG-00.
DR PDB; 1DK1; 03-AUG-00.
DR PDB; 1DK2; 03-AUG-00.
DR PDB; 1DK3; 03-AUG-00.
DR PDB; 1DK4; 03-AUG-00.
DR PDB; 1DK5; 03-AUG-00.
DR PDB; 1DK6; 03-AUG-00.
DR PDB; 1DK7; 03-AUG-00.
DR PDB; 1DK8; 03-AUG-00.
DR PDB; 1DK9; 03-AUG-00.
DR PDB; 1DKA; 03-AUG-00.
DR PDB; 1DKB; 03-AUG-00.
DR PDB; 1DKC; 03-AUG-00.
DR PDB; 1DKD; 03-AUG-00.
DR PDB; 1DKE; 03-AUG-00.
DR PDB; 1DKF; 03-AUG-00.
DR PDB; 1DKG; 03-AUG-00.
DR PDB; 1DKH; 03-AUG-00.
DR PDB; 1DKI; 03-AUG-00.
DR PDB; 1DKJ; 03-AUG-00.
DR PDB; 1DKK; 03-AUG-00.
DR PDB; 1DKL; 03-AUG-00.
DR PDB; 1DKM; 03-AUG-00.
DR PDB; 1DKN; 03-AUG-00.
DR PDB; 1DKO; 03-AUG-00.
DR PDB; 1DKP; 03-AUG-00.
DR PDB; 1DKQ; 03-AUG-00.
DR PDB; 1DKR; 03-AUG-00.
DR PDB; 1DKS; 03-AUG-00.
DR PDB; 1DKT; 03-AUG-00.
DR PDB; 1DKU; 03-AUG-00.
DR PDB; 1DKV; 03-AUG-00.
DR PDB; 1DKW; 03-AUG-00.
DR PDB; 1DKX; 03-AUG-00.
DR PDB; 1DKY; 03-AUG-00.
DR PDB; 1DKZ; 03-AUG-00.
DR PDB; 1DK1; 03-AUG-00.
DR PDB; 1DK2; 03-AUG-00.
DR PDB; 1DK3; 03-AUG-00.
DR PDB; 1DK4; 03-AUG-00.
DR PDB; 1DK5; 03-AUG-00.
DR PDB; 1DK6; 03-AUG-00.
DR PDB; 1DK7; 03-AUG-00.
DR PDB; 1DK8; 03-AUG-00.
DR PDB; 1DK9; 03-AUG-00.
DR PDB; 1DKA; 03-AUG-00.
DR PDB; 1DKB; 03-AUG-00.
DR PDB; 1DKC; 03-AUG-00.
DR PDB; 1DKD; 03-AUG-00.
DR PDB; 1DKE; 03-AUG-00.
DR PDB; 1DKF; 03-AUG-00.
DR PDB; 1DKG; 03-AUG-00.
DR PDB; 1DKH; 03-AUG-00.
DR PDB; 1DKI; 03-AUG-00.
DR PDB; 1DKJ; 03-AUG-00.
DR PDB; 1DKK; 03-AUG-00.
DR PDB; 1DKL; 03-AUG-00.
DR PDB; 1DKM; 03-AUG-00.
DR PDB; 1DKN; 03-AUG-00.
DR PDB; 1DKO; 03-AUG-00.
DR PDB; 1DKP; 03-AUG-00.
DR PDB; 1DKQ; 03-AUG-00.
DR PDB; 1DKR; 03-AUG-00.
DR PDB; 1DKS; 03-AUG-00.
DR PDB; 1DKT; 03-AUG-00.
DR PDB; 1DKU; 03-AUG-00.
DR PDB; 1DKV; 03-AUG-00.
DR PDB; 1DKW; 03-AUG-00.
DR PDB; 1DKX; 03-AUG-00.
DR PDB; 1DKY; 03-AUG-00.
DR PDB; 1DKZ; 03-AUG-00.
DR PDB; 1DK1; 03-AUG-00.
DR PDB; 1DK2; 03-AUG-00.
DR PDB; 1DK3; 03-AUG-00.
DR PDB; 1DK4; 03-AUG-00.
DR PDB; 1DK5; 03-AUG-00.
DR PDB; 1DK6; 03-AUG-00.
DR PDB; 1DK7; 03-AUG-00.
DR PDB; 1DK8; 03-AUG-00.
DR PDB; 1DK9; 03-AUG-00.
DR PDB; 1DKA; 03-AUG-00.
DR PDB; 1DKB; 03-AUG-00.
DR PDB; 1DKC; 03-AUG-00.
DR PDB; 1DKD; 03-AUG-00.
DR PDB; 1DKE; 03-AUG-00.
DR PDB; 1DKF; 03-AUG-00.
DR PDB; 1DKG; 03-AUG-00.
DR PDB; 1DKH; 03-AUG-00.
DR PDB; 1DKI; 03-AUG-00.
DR PDB; 1DKJ; 03-AUG-00.
DR PDB; 1DKK; 03-AUG-00.
DR PDB; 1DKL; 03-AUG-00.
DR PDB; 1DKM; 03-AUG-00.
DR PDB; 1DKN; 03-AUG-00.
DR PDB; 1DKO; 03-AUG-00.
DR PDB; 1DKP; 03-AUG-00.
DR PDB; 1DKQ; 03-AUG-00.
DR PDB; 1DKR; 03-AUG-00.
DR PDB; 1DKS; 03-AUG-00.
DR PDB; 1DKT; 03-AUG-00.
DR PDB; 1DKU; 03-AUG-00.
DR PDB; 1DKV; 03-AUG-00.
DR PDB; 1DKW; 03-AUG-00.
DR PDB; 1DKX; 03-AUG-00.
DR PDB; 1DKY; 03-AUG-00.
DR PDB; 1DKZ; 03-AUG-00.
DR PDB; 1DK1; 03-AUG-00.
DR PDB; 1DK2; 03-A

```
Query Match      42.7%; Score 939.5; DB 1; Length 432;
Best Local Similarity 47.9%; Pred. No. 1.1e-66;
Matches 196; Conservative 61; Mismatches 137; Indels 15; Gaps 7;

QY 13 LVLMGLAAITAPVA---APSGYTLERVLVSRHGVRSPTKOTQOLMNDVTPDKWPQW 69
DB 5 LIPFLSLIPLTPQSAFAQSEPE-LKLESVIVSRHGVRAPTKATQOLMNDVTPDAWTPW 63
QY 70 VKAGYLTTPRGAEVLTMGGFYDYSGLLA-AGCPAEGGVYAQADIDORTLTGQAF 128
DB 64 VKLGLWTRGGELIAYLGHYQRLVADGLLAKKCPQSGQVAILADVDERTRKTGEAFA 123
QY 129 DGVAPGCGLTVHQADLKKTDPLPHVETGVCKLDNAQTDKAIERLGGPDLTVSQRYAK 189
DB 124 AGLAPDCAITVHTQADTSSPDLFNLKTVGQLENDANVTDAILSRAGGSIADTGHRT 183
QY 189 PFAQMGDVLNFAAASPYCKSLQOQKTCDFAFHAEVNVNKEGTVLSGPLALSSTLGE 248
DB 184 AFRELRLVNPQSNLCKREKQDECSLTQALPSLKVSD--NVSLTGAVSLASMLTE 241
QY 249 IFLQNAQAMPEVAQRKGAENWVLSLHNAQFNLMAKTPYIARHKGTPLLOQIDTAL 308
DB 242 IFLQQAQMPGPGGRITDQSHWNTLLSLHNAQFYLLQRTPEVARSRAATPLDLIKTAL 301
QY 309 T---LQDAQCKPLISAQNVFLGHDHTNIAIAGMLGADWOLPEQDNTPEGGELVF 365
DB 302 TPHPQKQAYGVTLPTS-----VLFIAGDHTNLALGALNLWTLPEQDNTPEGGELVF 357
QY 366 ELWQNPNDHQRYAVYKMFYQTMQDLRQNAEKLDRKNPAGIISVAVAGCE 414
DB 358 ERWRRLSDNSQWIVQSLVFQTLQQRDKTPLSL-NTTPGEVKLTLAGCE 405

RESULT 2
AGP_ECOLI
ID AGP_ECOLI STANDARD; PRT; 413 AA.
AC P19926;
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glucose-1-phosphatase precursor (EC 3.1.3.10) (Glpase).
GN AGP OR B1002.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]_TaxID=562;
RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-31.
RC STRAIN=K12;
RX MEDLINE=90130318; PubMed=2153660;
RA Pradel E., Marck C., Boquet P.L.;
RT "Nucleotide sequence and transcriptional analysis of the Escherichia
EL coli agp gene encoding periplasmic acid glucose-1-phosphatase.";
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RN Science 277:1453-1474 (1997).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikenoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA
```

```
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
RN [4]
RP SEQUENCE OF 23-32.
RC STRAIN=K12 / ENG2;
RX MEDLINE=97443975; PubMed=9298646;
RA Link A.J., Robison K., Church G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
RT in the genome of Escherichia coli K-12.";
RL Electrophoresis 18:1259-1313(1997).
CC -1- PHOSPHATE ABSOLUTELY REQUIRED FOR THE GROWTH OF E. COLI IN A HIGH-
CC PHOSPHATE MEDIUM CONTAINING G-1-P AS THE SOLE CARBON SOURCE.
CC -1- CATALYTIC ACTIVITY: D-glucose 1-phosphate + H(2)O = D-glucose +
CC phosphate.
CC -1- ENZYME REGULATION: INDEPENDENT FROM INORGANIC PHOSPHATE
CC AVAILABILITY, AND APPARENTLY SUBMITTED TO CATABOLITE REPRESSION.
CC IT IS POSITIVELY CONTROLLED BY CAMP AND THE CAMP RECEPTOR PROTEIN.
CC -1- SUBUNIT: Homodimer.
CC -1- SUBCELLULAR LOCATION: Periplasmic.
CC -1- MISCELLANEOUS: OPTIMALLY ACTIVE AROUND PH 4.
CC -1- SIMILARITY: Belongs to the histidine acid phosphatase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M33807; AAA23426.1; --
CC EMBL; AS000202; AAC74087.1; --
CC EMBL; D90737; BAA35769.1; --
CC EMBL; D90738; BAA35779.1; --
CC PIR; JVO087; JVO087.
CC HSSP; P07102; 1DKM.
CC EcoGene; EGI0033; agp.
CC InterPro; IPR000560; HisAc_phsphtase.
CC Pfam; PF00328; acid_phosphat; 1.
CC PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
CC PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
CC Hydrolase; Periplasmic; Signal; Complete proteome.
FT SIGNAL 1 22
FT CHAIN 23 413 GLUCOSE-1-PHOSPHATASE
FT ACT_SITE 40 40 NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).
FT ACT_SITE 311 311 PROTON DONOR (BY SIMILARITY).
SQ SEQUENCE 413 AA; 45683 MW; ADADAD3639D0D6AB CRC64;

Query Match      24.1%; Score 530; DB 1; Length 413;
Best Local Similarity 31.3%; Pred. No. 2.3e-34;
Matches 131; Conservative 82; Mismatches 165; Indels 40; Gaps 11;

QY 10 LSGVLMLSLGLAAITAPVAAPSGYTLERVLVSRHGVRSPTKOTQOLMNDVTPDKWPQW 68
DB 10 VAGIVLLASNAQAQTV-----PEGYLQQLVLMMSRHLNRAPLANRSGVLEQSTPNKWPEN 64
QY 69 PVKAGYLTTPRGAEVLTMGGFYDYSGLLAAG-CPAEGGVYAQADIDORTLTGQAF 127
DB 65 DVPFGQLTTGGVLEVYMGHYNREWLAEQGMVKSCEPPYTVAYANSLQRTVATAQFF 124
QY 128 LDGVAPGCGLTVHQADLKKTDPLPHVETGVCKLDNAQTDKAIERLGGPDLTVSQRYA 187
DB 125 ITGAFPGCDIPVHQQKMGMDTFNPNVITDSDAAFSQAFAAMEKELS-----KLQUT 178
QY 188 KPPAQMGDVLNFAAASPYCKSLQOQKTCDFAFHAEVNVNKEGTVLSGPLALS 243
DB 179 DSIQLLEKLVNPKDSQPACKKEQCCSLVDGNT-----FSA-----KYQEPGVSGFLKVG 228
QY 244 STLGEIFLQNAQAMP--EVAWQRKGAENWVLSLHNAQFNLMAKTPYIARHKGTPL 301
DB
```

Db 229 NSLVDAFTLQYEGFPMQVAMGEIKSDQWKVLSKXGQYDLSFTSPVARNVAKPLV 288
 QY 302 QQIDTALTQLDQAGQKLPISAQNRVFLGGHDTNTANTAGMLG-ADWQLPEQPDNTPPG 360
 Db 289 SYIDKALVDR-----TSAPKTVLVGHDSNTASLTALDFPKYQLHQDQERTPIG 339
 QY 361 GGLVFFELWQPNHQRVAVKMFYQTMQDLNRNAEKLDLKNPNAGIISVAVAGCNGD 418
 Db 340 KGVIFQRWHDSSKANRLMKIEYVYQSAEQLRNADALTQ-APAQRVTLBLSGCPIDAD 396

RESULT 3
 AGP_SALTY STANDARD; PRT; 413 AA.
 AC O33921;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Glucose-1-phosphatase precursor (EC 3.1.3.10) (Gipase).
 GN AGP OR STM1117.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Lt2 / SGSC1412 / AFCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nian M.,
 RA Waterston R., Wilson R.K.;
 RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
 LT2.";
 RT Nature 413:852-856(2001).
 RL [2]
 RN SEQUENCE OF 1-249 FROM N.A.
 RP STRAIN=TN1379;
 RX MEDLINE=37405892; PubMed=9260936;
 RA Gupta S.D., Wu H.C., Rick P.D.;
 RA "A Salmonella typhimurium genetic locus which confers copper
 RT tolerance on copper-sensitive mutants of Escherichia coli.";
 RL J. Bacteriol. 179:4977-4984(1997).
 CC -1- CATALYTIC ACTIVITY: D-glucose 1-phosphate + H(2)O = D-glucose +
 CC phosphate.
 CC -1- SUBUNIT: Homodimer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Periplasmic.
 CC -1- SIMILARITY: Belongs to the histidine acid phosphatase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@sb-sib.ch).
 CC -----
 CC EMBL; AE008748; AAL20049.1; -
 CC EMBL; U75949; AAC45604.1; -
 CC HSSP; P07102; 1DKL.
 DR StyGene; SG10595; agp.
 DR InterPro; IPR000560; HisAc_pheptase.
 DR Pfam; PF00328; acid_phosphat; 1.
 DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
 DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
 KW Hydrolase; Periplasmic; Signal; Complete proteome.
 FT SIGNAL 1 22 BY SIMILARITY
 FT CHAIN 23 413 GLUCOSE-1-PHOSPHATASE.
 FT ACT_SITE 40 40 NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).
 FT ACT_SITE 311 311 PROTON DONOR (BY SIMILARITY).
 FT CONFLICT 57 57 T -> P (IN REF. 2).
 FT CONFLICT 60 63 AWA -> TCP (IN REF. 2).

FT CONFLICT 74 74 K -> T (IN REF. 2).
 SQ SEQUENCE 413 AA; 45559 MW; E057667CF8A7244C CRC64;
 Query Match 22.7%; Score 499; DB 1; Length 413;
 Best Local Similarity 32.0%; Pred.No. 6.4e-32;
 Matches 135; Conservative 77; Mismatches 166; Indels 44; Gaps 14;
 QY 8 VRLSGLVLMLSGLAAITAPVAAEPSGYTLRVRVILSRHVRSP-TKQTQLMNDVTPDKWP 66
 Db 8 VAVAGAVLLSSAVQAQTT-----PEGYLOQVLMMSRNLRAPLANNNGSVLAQSTPNWP 62
 QY 67 QWPKAGYLTFRGAELVTLMGGFYDIFRSIGLLAAG-CPAEGGVYAAQADIDQRTLTGQ 125
 Db 63 AWDVPGGQLTTKGGVLEVTYHGYTRVWLVAGQIFPSGECFAPDPTVYAYANSLQRTVATAQ 122
 QY 126 AFLDGVAPGCGLTVHNOADLKTKDPLFHPVETGVCKLDNAQTDKAIERLGGGLDVTYSQR 185
 Db 123 FFIIGAPGCDIPVHQEKMGWMDPTNPNVITDDSAFRQAVQAMEK-----ASSQL 175
 QY 186 YA-KPFAQMGDVLFNFAASPYCKSLQQQKCTCDFAH-----FAANEVNVNKGTKVTLGSP 240
 Db 176 HLDESYSKLEQITHYQDPSCKEKHQ-----CSLIDAKDTFSANYQQ-----EPGVQGPL 225
 QY 241 ALSSTLGEIFLLQNAQAMP--EVAWQRLKGAENWVSLSLHNAQFNLMAKTPYIARHGT 298
 Db 226 KVGNSLVDAFTLQYEGFPMQVAMGGIHTRQKVLKXGQYDLSFTSPVARNVAA 285
 QY 299 FLLOQIDTALTQLDQAGQKLPISAQNRVFLGGHDTNTANTAGMLG-ADWQLPEQPDNT 357
 Db 286 FLVKYIDKVLAD-----RVGAP-KVTVLVGHDSNTASLTALDFPKYQLHQDQERT 336
 QY 358 PPGGLVAFELWQPNHQRVAVKMFYQTMQDLNRNAEKLDLKNPNAGIISVAVAGC--EN 415
 Db 337 FIGQLVFRWHDGNARLDMKIEYVYQSAEQLRNADALTQ-SPAQRVTLBLSGCPVDA 395
 QY 416 NG 417
 Db 396 NG 397
 RESULT 4
 AGP_PRORE STANDARD; PRT; 417 AA.
 AC Q52309;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Glucose-1-phosphatase precursor (EC 3.1.3.10) (Gipase).
 GN AGP.
 OS Providencia rettgeri.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Providencia.
 OX NCBI_TaxID=587;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PV7;
 RA Riccio M.L., Chiesurin A., Lombardi G., Satta G.;
 RL Submitted (SEP-1992) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: D-glucose 1-phosphate + H(2)O = D-glucose +
 CC phosphate.
 CC -1- SUBUNIT: Homodimer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Periplasmic (By similarity).
 CC -1- SIMILARITY: Belongs to the histidine acid phosphatase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@sb-sib.ch).
 CC -----
 CC EMBL; X68201; CAA48288.1; -

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U00031; AAK18870.2; --
 DR HSP; P20646; 1RPA.
 DR WormPep; B0361.7; CE32100.
 DR InterPro; IPR000560; HisAc_phsphtse.
 DR Pfam; PF00328; acid_phosphat; 1.
 DR PROSITE; PS00616; HIS ACID PHOSPHAT_1; 1.
 DR PROSITE; PS00778; HIS ACID PHOSPHAT_2; 1.
 DR PROSITE; PS00778; HIS ACID PHOSPHAT_2; 1.
 KW Hypothetical protein; Signal; Glycoprotein; Hydrolase.
 FT SIGNAL 1 13 POTENTIAL.
 FT CHAIN 14 422 PUTATIVE ACID PHOSPHATASE B0361.7.
 FT ACT_SITE 40 40 NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).
 FT ACT_SITE 80 80 BY SIMILARITY.
 FT ACT_SITE 278 278 PROTON DONOR (BY SIMILARITY).
 FT DISULFID 152 363 BY SIMILARITY.
 FT DISULFID 205 302 BY SIMILARITY.
 FT DISULFID 338 342 BY SIMILARITY.
 FT CARBOHYD 104 104 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 210 210 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 218 218 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 312 312 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 323 323 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 422 AA; 48408 MW; 3E29FACB4FAE0CAC CRC64;
 Query Match 5.7%; Score 125; DB 1; Length 422;
 Best Local Similarity 20.5%; Pred.No. 0.022;
 Matches 91; Conservative 58; Mismatches 161; Indels 134; Gaps 20;
 QY 13 LVLMISGLAAITA-----PVAAPSGYTLERVILSRHGVRSPTKOTLMNDVTPDKWP 66
 DB 4 LVLVLLGASGINAVVYKEVPIQANTD--TLEYVHTVWRHGDRTPAEL-----LFPDDIT 55
 QY 67 QWPVKAGYLTPRGAE-----LVILMGGFYGDYFSLGLLAAGCPAEGGVYAAQADIDQ 118
 DB 56 KWPEGLGELTEQGAQAQYRLGQWLKRYGSLWKEFN-----RNAIIRSDYN 104
 QY 119 RTRITGQAFLDGVAP-----CGGLTVHNOADKKTDPLFHPVETGVCKLDNAOTDKAIEE 173
 DB 105 RTLMSAQAQWAGLFPFKYPIAGGL-----MWQPI--PVHT-----ISKPTDKELYE 148
 QY 174 RLGGPLDTVQRYAKPPAQMGDVLPFAAPYCKSLQOQKTCDFAH---PAAENVVYKE 230
 DB 149 EASCTPAEIMN-----AOW-----KSTRANGIRKKFARELSFFSOKLNPMM 191
 QY 231 GTKVTLSGLALSLTGLGEIFLLQNAQAMPEVAVQRLKGAENWVSLSLHNAQFNLMAKTP 290
 DB 192 ELKAT-----WRIFDNLCEKQNNITPWSMNSSIFERVDQLYNEVSQLEPHTD 240
 QY 291 YIARHKGTPLLQIDTALTQLDQAQOKLPISQNRVFLGGHDTNTANTAGMLGADWQL 350
 DB 241 TLRELGRGTLLLEIFHRFSK--ASGS---LGKEAKFYASAHDSSTIAALLATLIGVY-- 293
 QY 351 PEQPDTPPGGLVPFELWQPNHQRVYAVKMFQVTMDQLRNAEKL-----DLKNN 401
 DB 294 -----DIYPKATCLLI--EMHKLANTLIRVPHKNETDIDR- 329
 QY 402 PAGIISVAVAGCENN-----GDD 419
 DB 330 ---LIEYSIPGCDPCTQLKLGDD 350
 RESULT 6
 ID PPAL_MOUSE
 AC P24638; O8QZT5;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)

DR PIR; S25627; S25627.
 DR HSP; S07102; 1DXL.
 DR InterPro; IPR000560; HisAc_phsphtse.
 DR Pfam; PF00328; acid_phosphat; 1.
 DR PROSITE; PS00616; HIS ACID PHOSPHAT_1; 1.
 DR PROSITE; PS00778; HIS ACID PHOSPHAT_2; 1.
 KW Hydrolase; Periplasmic; Signal.
 FT SIGNAL 1 23 POTENTIAL.
 FT CHAIN 24 417 GLUCOSE-1-PHOSPHATASE.
 FT ACT_SITE 42 42 NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).
 FT ACT_SITE 314 314 PROTON DONOR (BY SIMILARITY).
 SQ SEQUENCE 417 AA; 46940 MW; 8CC4CA4F5DBFB90 CRC64;
 Query Match 21.0%; Score 462; DB 1; Length 417;
 Best Local Similarity 30.6%; Pred.No. 5.4e-29;
 Matches 128; Conservative 84; Mismatches 170; Indels 36; Gaps 15;
 QY 14 VLMLISGLAAITAPVAA-----EPGTYTLERVILSRHGVRSPTKOTLMNDVTPDKWP 67
 DB 5 VLTLCLSAALFAPIAPTWASTDNQADMVLDQVLVLSRHLRTPIVNTGILTEVTDKKWPD 64
 QY 68 WPKVAGYLTPRGAEVTLMGFGVGYFSLGLLAAG-CPAEG-GVYAAQADIDQRTLTGQ 125
 DB 65 WDAKSGYLTTQGGALEVYGHYFREWIDQNLKADDELCTNSNEDILYTNLSLQRTIAQA 124
 QY 126 AFLDGVAPGGLTVHNOADKKTDPLFHPVETGVCKLDNAQDKAIEERLGGPLDVTVSQR 185
 DB 125 PFAAGAFPOCKVNIHQPEIGKWDVFNPIITNGSPFEKQKALAAAMDYLLKG----LSLK 180
 QY 186 YAKPAQMGDVLPFAAPYCKSLQOQKTCDF--AHPAANEVNVNKEGTYKVTLSGLALSS 244
 DB 181 AG--YEELDTVLNKDSQCKT-----DKLCLNDSQKNSFIIEADKE---PGVSGPLKIAN 231
 QY 245 TLGEIFLLQNAQAMP--EYAWORLKAENWVSLSLHNAQFNLMAKTPYIARHKGTPLLQ 302
 DB 232 SAVDAIDLQYEGFPADQVANGLVDPPEKKWKLNTLKNAYQETLTPPKIAKNVAHPILN 291
 QY 303 QIDTALTQLDQAQOKLPISAQNRVFLGGHDTNTANI--AGMLGADWQLPEQPDTPPGG 361
 DB 292 YIDKGF-VSDV-KGETA-----KFILVGHDSNLSIASMSAMDFKPYQLAQOYEHTPIGG 343
 QY 362 GLVFELWQPNHQRVYAVKMFQVTMDQLRNAEKLDKNNPAGIISVAVAGC--ENNG 417
 DB 344 KLVFQRTWKQTKQFMKVEYVYQADQLRDNAYLSLETPPKH-VTLELKCDCPVKNG 400
 RESULT 5
 ID PPAY_CABEL
 AC Q10974;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Putative acid phosphatase B0361.7 precursor (EC 3.1.3.2).
 GN B0361.7.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Paloderinae; Caenorhabditis.
 OC NCBI_TaxID=6239;
 RX [1] -----
 RN SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Du Z.;
 RL Submitted (FEB-1994) to the EMBL/GenBank/DBJ databases.
 RN REVISIONS.
 RA Waterston R.;
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 CC -/- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O = an
 CC alcohol + phosphate.
 CC -/- SIMILARITY: Belongs to the histidine acid phosphatase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

10-OCT-2003 (Rel. 42, Last annotation update)
Lysosomal acid phosphatase precursor (EC 3.1.3.2) (IAP).
ACP2.
Mus musculus (Mouse).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
TISSUE=Liver;
MEDLINE=22388257; PubMed=12477932;
Srausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
Raha S.S., Lomellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
Boak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyx S.W.,
Vallalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahy J., Hulton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalusz D.E.,
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
"Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[2]
SEQUENCE OF 3-423 FROM N.A.
MEDLINE=91282985; PubMed=2059337;
Geier C., von Figura K., Pohlmann R.;
"Molecular cloning of the mouse lysosomal acid phosphatase."
Biol. Chem. Hoppe-Seyler 372:301-304(1991).
CC -!- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O = an
alcohol + phosphate.
CC -!- SUBCELLULAR LOCATION: Lysosomal.
CC -!- SIMILARITY: Belongs to the histidine acid phosphatase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; BC023343; AAH23343.1; -;
EMBL; X57199; CA40485.1; -;
PIR; S14742; S14742.
DR HSSP; P15309; 2HPA.
DR MGD; MGI:87882; Acp2.
DR InterPro; IPR000560; HisAc_phosphatase.
DR Pfam; PF00328; acid_phosphat; 1.
DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
KW Hydrolase; Signal; Glycoprotein; Lysosome.
FT SIGNAL 1 30
FT CHAIN 31 423 LYSOSOMAL ACID PHOSPHATASE.
FT DISULFID 159 370 BY SIMILARITY.
FT DISULFID 212 310 BY SIMILARITY.
FT DISULFID 345 349 BY SIMILARITY.
FT ACT SITE 42 42 NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).
FT ACT SITE 84 84 PROTON DONOR (BY SIMILARITY).
FT ACT SITE 286 286 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 92 133 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 167 167 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 177 177 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 197 197 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 267 267 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 374 374 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 131 131 N -> S (IN REF. 2).
SQ SEQUENCE 423 AA; 48508 MW; AFD7ASC90AF2AF4 CRC64;
Query Match 5.6%; Score 124; DB 1; Length 423;
Best Local Similarity 22.8%; Pred. No. 0.027;
Matches 100; Conservative 52; Mismatches 148; Indels 138; Gaps 25;
QY 11 SGLVLMGLA-AITAPVAEPEGYTLERVVILSRHGVRSPTKTQTLMDVTPDKWPWP 69
DB 11 AALLQFLMGLMCLTVPFIQAR-----SURFTLLYRHGDRSPVK-TYFKDPYQBEKWFQ-- 63
QY 70 VKAGYLTPRGA-----EL-VTLMGFFYG----DYFRSLGLLAAGPAGGGVYQAADIQORT 120
DB 64 -GFQGLTKEGMLQHWELQALRQRYHGLNLSVHRQ-----EYVVRSTDFDRT 110
QY 121 RLTSQAFLDGVARGCGLTVHQADLKKTDPLFHFVEVGVCKLDNAQTKAIEERLGGPLD 180
DB 111 LMSAEANLAGLFF-----PNEVOHFNPNISWQDIP-----VHTVPTEDRL----- 151
QY 181 TVSQRYAKPFAQMGDVLNFAASPYCKSLQOQKGT-----CDFAHFAANVNV 227
DB 152 -----LKFPLGCPRYEQNLQNETRQTPYQNRSIQNAQFLNVMVANETGL 195
QY 228 NKEGTVKTLISPLALSSTLGEIP-----LLQNAQAMPEVAMORLKGAE--NVVSL 275
DB 196 -----TNVTL-----ETIMNVYDTLCEQTHGLLLPFWASPTV-QLSLQDKFSLFL 243
QY 276 LSLHNAQFNLMAKTPYIARHKGSTPLLOQIDTALTQLDAGQKLPISAQNRVFLGGHDT 335
DB 244 FGIH-----EQVQKARLQGVLLAQILKNLTLM--ATTSQFP-----KLLVYSADHT 288
QY 336 NIANIAGMLGADWQLPEQDPTPPGGGLVFEELQNPNDHQRYVAVKMFYQTMQDLRNAREK 395
DB 289 TLVALQMALNV-----YNGKQAPYASCHIFELYQEDNGN----FSVEMTF-----RN--- 331
QY 396 LDLKNNPAGIISVAVAGC 413
DB 332 -DSKKAPWPLI---LPGC 345
RESULT 7
PPAL HUMAN
ID PPAL HUMAN STANDARD; PRT; 423 AA.
AC P11117; Q9BTU7;
DT 01-JUL-1989 (Rel. 11, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Lysosomal acid phosphatase precursor (EC 3.1.3.2) (LAP).
GN ACP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RC SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Placenta;
RX MEDLINE=89052645; PubMed=3191910;
RX Pohlmann R., Krentler C., Schmidt B., Schroeder W., Lorkowski G.,
Culley J., Mersmann G., Geier C., Waheed A., Gottschalk S.,
Grzeschik K.H., Hasikik A., von Figura K.;
"Human lysosomal acid phosphatase: cloning, expression and
chromosomal assignment."
EMBO J. 7:2343-2350(1988).
[2]
RN SEQUENCE FROM N.A.
RP TISSUE=Leukocyte;
RX MEDLINE=89177428; PubMed=2776754;
RX Geier C., von Figura K., Pohlmann R.;

"Structure of the human lysosomal acid phosphatase gene.";
Eur. J. Biochem. 183:611-616(1989).
[3]
SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Boeak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Rulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettner M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Rodriguez A.C., Trinchwood J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Trinchwood J., Schmutz J., Myers R.M.,
RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smalusz D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O = an
alcohol + phosphate.
CC -!- SUBCELLULAR LOCATION: Lysosomal.
CC -!- SIMILARITY: Belongs to the histidine acid phosphatase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; X12548; CAA31064.1; -;
DR EMBL; X15525; CAA33542.1; -;
DR EMBL; X15526; CAA33542.1; JOINED.
DR EMBL; X15527; CAA33542.1; JOINED.
DR EMBL; X15528; CAA33542.1; JOINED.
DR EMBL; X15529; CAA33542.1; JOINED.
DR EMBL; X15530; CAA33542.1; JOINED.
DR EMBL; X15531; CAA33542.1; JOINED.
DR EMBL; X15532; CAA33542.1; JOINED.
DR EMBL; X15533; CAA33542.1; JOINED.
DR EMBL; X15534; CAA33542.1; JOINED.
DR EMBL; X15535; CAA33542.1; JOINED.
DR EMBL; SC003160; AA03160.1; -;
DR PIR; S06167; S06167.
DR HSP; P20646; IRPA.
DR Genew; HGNC:123; ACP2.
DR MIM; 171650; -;
DR GO; GO:0016021; C:integral to membrane; TAS.
DR GO; GO:0003993; F:acid phosphatase activity; TAS.
DR InterPro; IPR000560; HisAc_Phosphatase.
DR Pfam; PF00328; acid_phosphat; 1.
DR PROSITE; PS00216; HIS ACID PHOSPHAT 1; 1.
DR PROSITE; PS00778; HIS ACID PHOSPHAT 2; 1.
KW Hydrolase; Signal; Glycoprotein; Lysosome.
FT SIGNAL 1 30
FT CHAIN 31 423 LYSOSOMAL ACID PHOSPHATASE.
FT DISULFID 159 370 BY SIMILARITY.
FT DISULFID 212 310 BY SIMILARITY.
FT DISULFID 345 349 BY SIMILARITY.
FT ACT_SITE 42 42 NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).
FT ACT_SITE 84 84 BY SIMILARITY.
FT ACT_SITE 286 286 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 92 92 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT

FT CARBOHYD 167 167 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 177 177 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 267 267 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 29 29 Q -> R (IN REF. 1 AND 2).
SQ SEQUENCE 423 AA; 48316 MW; 1A2CBF40ED3724B CRC64;
Query Match 5.5%; Score 121.5; DB 1; Length 423;
Best Local Similarity 22.2%; Pred. No. 0.042;
Matches 103; Conservative 51; Mismatches 164; Indels 145; Gaps 23;
QY 11 SGVLMLSLGLAATAPVAEPGTYLLEVRVILSRHGVSPKOTQQLMNDVTPDKWP- 66
DB 11 AALLQLLGLNVLVMPPTQARS---LRFTLLYRHGDRSPVK---TTPKDYQSE 59
QY 67 QNPVKGAGYTPRGA---EL-VTLMGFFYG---DYFRSLGLLAAGCPAEGGYAAQADID 117
DB 60 ENPQFGQLTKGMLQHWELGQALRQRYHGLNTSYHQ- -EYVYRSTDF 107
QY 118 QRTFLTGQAFLDGAVPGCLTVHQADLKTDPHPVETGVCKLDNAQTDKAIERLGG 177
DB 108 DRTLMSAEANLAGLFPFGMGQRFN-----PNISWQFIP-----VHTVPITEDRL-- 151
QY 178 PLDTVSQRYAKPFAQMGDVNLNFAASPYCKSLQQQKGT-----CDFAHPAANE 224
DB 152 -----LKPGLGPCRYEQQLNETROTPEYQNESSRNAQFLDMVANE 192
QY 225 VVNKKGKTVLSGPIALSTTGEITLQNAQAMPEVAQWLKGAENWVLSLHNAQFN 284
DB 193 TGL-----TDLTLETWNVYDYL-----FCQTHGLRLPPW---ASPQTMQRLSLKDFSPR 241
QY 285 LM---AKTPVIARHKGTPLLOQIDTALTQLDAQGGKLPISAGNRVFLFGHDTNANIA 341
DB 242 FLGIYQQAEKARLQGVVLAQIRKNTLM--ATTSQLP-----KLIVYSAHDTTL--VA 292
QY 342 GMLGADWQLPEQPDNTPPGGGLVFLWQNPNDHQRVYAVVMFYQT----- 386
DB 293 LQALDVTNGEQ---APYASCHIFELYQSDSGN---FSVEMYFRNESDKAPWLSLPGCP 346
QY 387 -----MDQLRNAR-----KLDLKNNPAG---IISVAVAG 412
DB 347 HRCPLQDFLRLETPVVKDWQEQCLASGPADEVIVALLAVCG 389
RESULT 8
ID NU57 YEAST STANDARD; PRT; 541 AA.
AC P48837;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Nucleoporin NUP57 (Nuclear pore protein NUP57).
GN NUP57 OR YGR119C OR G6320.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95129554; PubMed=7828598;
RA Grandi P., Schlaich N., Tekotte H., Hurt E.C.;
RT "Functional interaction of Nup57 with a core nucleoporin complex
consisting of Nup133, Nup49p and a novel protein Nup57p.";
RL EMBL J. 14:76-87(1995).
RN [2]
RP SEQUENCE OF 1-354 FROM N.A.
RX STRAIN=S288C / FY1679;
RX MEDLINE=97157982; PubMed=9046098;
RA van Dyck L., Tettelin H., Purnelle B., Goffeau A.;
RT "An 18.3 kb DNA fragment from yeast chromosome VII carries four
unknown open reading frames, the gene for an Asn synthase, remnants

```
RT of Ty and three trna genes." ;
RL Yeast 13:171-176 (1997).
RN [3]
RP SEQUENCE OF 243-541 FROM N.A.
RX MEDLINE=97061913; PubMed=8905931;
RA Hansen M., Albers M., Backes U., Coblenz A., Leuther H., Neu R.,
RA Schreier A., Schaefer B., Zimmermann M., Wolf K.;
RT "The sequence of a 23.4 kb segment on the right arm of chromosome VII
RT from Saccharomyces cerevisiae reveals CLB6, SPT6, RP29A and NUP57
RT genes, a Ty3 element and 11 new open reading frames." ;
RL Yeast 12:1273-1277 (1996).
CC -1- FUNCTION: Part of the nucleoporin complex; required for protein
CC transport in the nucleus.
CC -1- SUBUNIT: Interacts with NSP1, NUP49 and NUP96.
CC -1- SUBCELLULAR LOCATION: Nuclear pore complex.
CC -1- DOMAIN: Contains G-L-F-G repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X81155; CAA57053.1; -
DR EMBL; X83095; CAA58153.1; -
DR EMBL; Z72904; CAA97129.1; -
DR EMBL; Z72905; CAA97131.1; -
DR PIR; S51799; S51799.
DR Germline; 141431; -
DR SGD; S0003351; NUP57.
DR GO; GO:0005643; C:nuclear pore; IMP.
DR GO; GO:0006406; P:nuclear-nucleus import; IMP.
DR GO; GO:0006606; P:protein-nucleus import; IMP.
KW Nuclear protein; Transport; Coiled coil; Repeat.
FT DOMAIN 76 223 9 X 4 AA REPEATS OF G-L-F-G.
FT DOMAIN 26 31 POLY-ASN.
FT DOMAIN 127 130 POLY-THR.
FT DOMAIN 217 220 POLY-GLY.
FT DOMAIN 258 266 POLY-GLN.
FT DOMAIN 277 280 POLY-GLN.
FT DOMAIN 398 425 COILED COIL (POTENTIAL).
FT DOMAIN 541 AA; 57498 MW; B292ADF7B1D7B83C CRC64;
SQ SEQUENCE 541 AA; 57498 MW; B292ADF7B1D7B83C CRC64;

Query Match 4.8%; Score 106.5; DB 1; Length 541;
Best Local Similarity 23.8%; Pred. No. 0.89;
Matches 76; Conservative 41; Mismatches 120; Indels 83; Gaps 18;

Qy 77 PRGAELVTLMGGYDFRSLGLLAAGCPAEGGVYAQADIDQTRITGQAFLDGVAPGCG 136
Db 196 PQGS---TTTGGFLGSGTQNNLTGG-----GGLFGSQSQPQ-----TNTAPGLG 237
Qy 137 LTVENQADLKTDPLFHPVETGVCKLDNAQTDKAIERLGGPLDVTYSQYAKPFAQMGDV 196
Db 238 NTVSTQSPFAWKP-----STG-----SNLQOQQ--QQIQVPL-----QQTQALAQOQL 281
Qy 197 LNFAASPYCKSLQOQKTC-----DFAHFAANVNVNKGKTVTLGSLALST 245
Db 282 SN-----YPOQIQOVLKCKESWDPNNTTKTLRAFYVNKNV-----ETEALYTKP----- 327
Qy 246 LGEIFLLQNAQM-----PEVAMORLKGAEWWSLLSLHNAQFNLMKTPYIARH---KG 297
Db 328 -GHVLQEWQAEKKEKSPQIPIQIYGF-----GLNQNVQVTENVAQARIILNHILEKS 383
Qy 298 TPLIQ--QIDTALTQLDAQOQKLPISAQNRVFLGCHDTNINIAIAGMLG-----ADM 348
Db 384 TOLQKQHELDASRI-LKAQSRNVEI--EKRLKLGTLQATLAKNRLPLGIAEKKWSQF 440
Qy 349 QLPQPDNTPGGGLVFELW 368
Db 441 QTLQRSEDPAGLKTNELW 460
```

```
RESULT 9
PPAL_RAT PPAL_RAT STANDARD; PRT; 423 AA.
AC P20611;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Lysosomal acid phosphatase precursor (EC 3.1.3.2) (LAP).
GN AC2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89350910; PubMed=2764916;
RA Himeno M., Fujita H., Noguchi Y., Kono A., Kato K.;
RT "Isolation and sequencing of a cDNA clone encoding acid phosphatase
RT in rat liver lysosomes." ;
RL Biochem. Biophys. Res. Commun. 162:1044-1053 (1989).
CC -1- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O = an
CC alcohol + phosphate.
CC -1- SUBCELLULAR LOCATION: Lysosomal.
CC -1- SIMILARITY: Belongs to the histidine acid phosphatase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M27893; AAA0744.1; -
DR PIR; A33395; A33395.
DR HSRP; P20646; LRPA.
DR InterPro; IPR000560; HisAc phosphatse.
DR Pfam; PF00328; acid phosphat; 1.
DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
KW Hydrolase; Signal; Glycoprotein; Lysosome.
FT SIGNAL 1 30
FT CHAIN 31 423 LYSOSOMAL ACID PHOSPHATASE.
FT DISULFID 159 370 BY SIMILARITY.
FT DISULFID 212 310 BY SIMILARITY.
FT DISULFID 345 349 BY SIMILARITY.
FT ACT_SITE 42 42 NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).
FT ACT_SITE 84 84 BY SIMILARITY.
FT ACT_SITE 286 286 PROTON DONOR (BY SIMILARITY).
FT CARBOHYD 92 92 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 133 133 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 167 167 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 177 177 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 191 191 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 197 197 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 267 267 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 322 322 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 331 331 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 423 AA; 48319 MW; 6F4C4819CADD4496 CRC64;

Query Match 4.8%; Score 106; DB 1; Length 423;
Best Local Similarity 23.6%; Pred. No. 0.71;
Matches 94; Conservative 49; Mismatches 149; Indels 106; Gaps 23;

Qy 10 LSGVLMLSGMAAITAPVAAPSPGYTLERVVILSRHGVRSPTKQTQLMNDVTPDKWPMP 69
Db 17 LLGMCLM-----VMPPIQAR-----SLRFVTLLYRHGDRSPVK-AYPKDPYQBEKWPQ-- 63
Qy 70 VKAGVLTFRGAELVTLMGGFYGDYFRSL--GLLAAGCPAEGGVYAQADIDQTRITGQAF 127
Db 64 -GFGQLTKEG-----MLQHWELGQALRORYHGFLNASTHRQ--EVTVRSTDFTDLMSAEAN 117
```

```
QY 128 LDGAPGGLTVHQAADLKKTDP--LFHPVETGVCKLDNAQTDKAIBERLGGGLDVTQSQR 185
DB 118 LAGLFP-----PTEVQHFNPNISWQPIP-----VHTVPIEDRL----- 151
QY 186 YAKPFAQGVNINPAASPYCKSLQOQKKT-----CDFAHFAANE-----VNVNK 229
DB 152 -----LKFFLGCPCERYEQLONETRQTPEYQNMISQNAQFLDMVANETGLMNLTL 200
QY 230 EGTKVTLGGPLALSSTLGEIFLLQNAQAMPV--AWORLKGAEWVSLLSLHNAQFLMA 287
DB 201 E-TWNYVDYTLFCQTHG--LLLPFWASQTVQALSQKDF-SFLFLGFIHQD-----VQ 251
QY 288 KTPYIAHKGPFLQOQIDTALTOLDACQKQLP:SAQNRVFLFGGHTNTIANIAGMLGAD 347
DB 252 K-----ARLQGVLLAQILKNTLM--ATTSQFP-----KLVYSAHDTTLVALQMALNV- 299
QY 348 WQLPEQPDNTPGGGLVFLWQNDHQRYVAVKMFYQ 385
DB 300 -----YNGKQAPYASCHIFELVQEDNGN-----PSVENYFR 330

RESULT 10
VL1 CRPVK
ID VL1 CRPVK STANDARD; PRT; 505 AA.
AC P03102;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Major capsid protein L1.
GN L1.
OS Cotton-tail rabbit (shope) papillomavirus (strain Kansas) (CRPV).
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OC NCBI_TaxID=31553;
RX [1]
SEQUENCE FROM N.A.
MEDLINE=85166175; PubMed=2984661;
RA Giri I., Danos O., Yaniv M.;
RL "Genomic structure of the cottontail rabbit (Shope) papillomavirus.";
Proc. Natl. Acad. Sci. U.S.A. 82:1580-1584(1985).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; K02708; -; NOT ANNOTATED_CDS.
DR PIR; A03643; PIWLRB.
DR InterPro; IPR002210; PV_capsid_L1.
DR InterPro; IPR008975; Viral_cap_coat.
DR Pfam; PF00500; late_protein_L1; 1.
DR PRINTS; PR00865; HPVCAPSID.L1.
DR ProDom; PD000544; PV_capsid_L1; 1.
KW Coat protein; Late protein.
SQ SEQUENCE 505 AA; 57933 MW; 50FF168D8A2F5A38 CRC64;

Query Match
Best Local Similarity 4.7%; Score 103.5; DB 1; Length 505;
Matches 79; Conservative 42; Mismatches 118; Indels 113; Gaps 21;

QY 92 DYRSLG-----LLAAGC-PAEGGVYAA-----DIDQRT-----RLTGOAFLDG--V 131
DB 142 DYRKSMAFPKQVQLLMGLGCVATGEHWAQKCAEDPPQDPCPIELVNTVIEDGDMC 201
QY 132 APCGGLTVN--QADLKKTDPLEHPVETGVCKLDNAQTDKAIBERLGGGLDVTQSQR---Y 186
DB 202 EIGFGAMDKHTLOAS:SEV-----PLEAQISKYPDYLMQKQDGSDFYARREQY 256
QY 187 AKPFAQM--GDVLNFAASPYCKSLQOQKKTCDFAHFAANEVNVNKGTVTLGSLALSS 244
```

```
DB 257 ARHFFSRAGGDKENYKSRAYIKRTOMQGEA-----NANIATDNYCITPSSGLVSSD 307
QY 245 T--LGEIFLLQNAQAMPE--VAVORLKGAEWVSLLSLHNAQFLMA----- 287
DB 308 SQVFNRYWYLQNAQWNGVCH-----DNQIFVTVDNTRGTILSLVTKSKEQIKKTHG 361
QY 288 KTPYTA--RHKGTPLLQOIDTALTOLDACQKQLP:SAQNRVFLFGGHTNTIANIAGWL 344
DB 362 KTVHFFSYLRH-----VEEYELQFVLQ-----CKVKLTPENLSYLHSHRPTIIDN----- 407
QY 345 GADWOL-----PEQPDNTPGGGLVFLWQ 369
DB 408 ---WQLSVSAQSPGSLVDQRYLQSIATKCPPEPPKENTIDYKN--YKFEW 454

RESULT 11
6PGD CHLPN
ID 6PGD CHLPN STANDARD; PRT; 479 AA.
AC Q928T3; Q9JQC1;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 6-phosphogluconate dehydrogenase, decarboxylating (EC 1.1.1.44).
GN GND OR CPN0360 OR CP0398 OR CPB0369.
OS Chlamydia pneumoniae (Chlamydia) pneumoniae.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OC NCBI_TaxID=83558;
RX [1]
SEQUENCE FROM N.A.
STRAIN=CWL029;
MEDLINE=99206606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
RX [2]
SEQUENCE FROM N.A.
STRAIN=AR39;
MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
RX [3]
SEQUENCE FROM N.A.
STRAIN=J138;
MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
RX [4]
SEQUENCE FROM N.A.
STRAIN=TW-183;
RA Geng M.M., Schumacher A., Muehldorfer I., Bensch K.W., Schaefer K.P.,
RA Schneider S., Pohl T., Essig A., Marre R., Melchers K.;
RT "The genome sequence of Chlamydia pneumoniae TW183 and comparison with
RT other Chlamydia strains based on whole genome sequence analysis.";
RL Submitted (May-2003) to the EMBL/GenBank/DBJ databases
CC -1- CATALYTIC ACTIVITY: 6-phospho-D-glucuronate + NADP(+) = D-ribulose
CC 5-phosphate + CO(2) + NADPH.
CC -1- PATHWAY: Hexose monophosphate shunt.
CC -1- SIMILARITY: Belongs to the 6-phosphogluconate dehydrogenase
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```


Db 496 IVLKTQALCSENTDCIRLKNA-----LVNLGNKQWSALVKRAQSGNLEGMVLLK----- 544
 QY 308 LTIQLDAQCQKLPISAQNRVLFLEGGHTNIA--GMGLADWQLPEQPDTPFGGGLVF 365

QY 308 LTLQLDAQGQKLPISAQNRVFLFGHDTNIAIA--GMLGADWQLPEQPDNTPPGGGLVF 365

Db 547 -----PISAD-VL-----ENLINTAASSFYRETHLATEALNSPPGGFLI 586

Qy 366 ELWQNPD-NHQRVAVKMFYQTMOLNVAEK-----IDLKNNPAGIIS 407

Db 587 TSDEGKQLVNHPTLPLFDYSALSCWRELQRLSALLDTPFKAGIIT 635

RESULT 13

DYHC NEUCR

ID DYHC NEUCR STANDARD; PRT; 4367 AA.

AC P45443;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Dynein heavy chain, cytosolic (DYHC)

GN RO-1.

OS Neurospora crassa.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.

OX NCBI_TaxID=5141;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=74-OR23-1A / FGSC 987;

RX MEDLINE=95014704; PubMed=7929559;

RA Planann M.; Minke P.F.; Tinsley J.H.; Bruno K.S.;

RT "Cytosolic dynein and actin-related protein Arp1 are required for normal nuclear distribution in filamentous fungi.";

RL J. Cell Biol. 127:139-149(1994).

CC -!- FUNCTION: Cytosolic dynein acts as a motor for the intracellular retrograde motility of vesicles and organelles along microtubules. Dynein has ATPase activity; the force-producing power stroke is thought to occur on release of ADP. Required to maintain uniform nuclear distribution in hyphae.

CC -!- SUBUNIT: Consists of at least two heavy chains and a number of intermediate and light chains.

CC -!- SUBCELLULAR LOCATION: Cytoplasmic.

CC -!- DOMAIN: Dynein heavy chains probably consist of an N-terminal stem (which binds cargo and interacts with other dynein components), and the head or motor domain. The motor contains six tandemly-linked AAA domains in the head, which form a ring. A stalk-like structure (formed by two of the coiled coil domains) protrudes between AAA 4 and AAA 5 and terminates in a microtubule-binding site. A seventh domain may also contribute to this ring; it is not clear whether the N-terminus or the C-terminus forms this extra domain. There are four well-conserved and two non-conserved ATPase sites, one per AAA domain. Probably only one of these (within AAA 1) actually hydrolyzes ATP, the others may serve a regulatory function.

CC -!- SIMILARITY: Belongs to the dynein heavy chain family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL, L31504; AAA64908.1; --

DR PIR, B54802; B54802.

DR InterPro, IPR003593; AAA ATPase.

DR InterPro, IPR004273; Dynein heavy.

DR Pfam, PF03028; Dynein heavy; 1.

DR SMART, SM00382; AAA; 3.

KW Motor protein; Dynein; Microtubule; ATP-binding; Repeat; Coiled coil.

FT DOMAIN 1 1904 STEM (BY SIMILARITY)

FT DOMAIN 1905 2130 AAA 1 (BY SIMILARITY)

FT DOMAIN 2202 2460 AAA 2 (BY SIMILARITY)

FT DOMAIN 2566 2815 AAA 3 (BY SIMILARITY)

FT DOMAIN 2909 3179 AAA 4 (BY SIMILARITY)

FT DOMAIN 3193 3481 STALK (BY SIMILARITY)

FT DOMAIN 3565 3794 AAA 5 (BY SIMILARITY)

FT DOMAIN 4003 4215 AAA 6 (BY SIMILARITY).

FT DOMAIN 676 693 COILED COIL (POTENTIAL).

FT DOMAIN 1176 1215 COILED COIL (POTENTIAL).

FT DOMAIN 1327 1351 COILED COIL (POTENTIAL).

FT DOMAIN 1557 1574 COILED COIL (POTENTIAL).

FT DOMAIN 1637 1668 COILED COIL (POTENTIAL).

FT DOMAIN 2195 2218 COILED COIL (POTENTIAL).

FT DOMAIN 3193 3296 COILED COIL (POTENTIAL).

FT DOMAIN 3423 3481 COILED COIL (POTENTIAL).

FT DOMAIN 3778 3809 COILED COIL (POTENTIAL).

FT NP_BIND 1943 1950 ATP (POTENTIAL).

FT NP_BIND 2240 2247 ATP (POTENTIAL).

FT NP_BIND 2605 2612 ATP (POTENTIAL).

FT NP_BIND 2947 2954 ATP (POTENTIAL).

SQ SEQUENCE 4367 AA; 495568 MW; 1B10F3E2D170D6F CRC64;

Query Match 4.5%; Score 99.5; DB 1; Length 4367;

Best Local Similarity 18.0%; Pred. No. 49;

Matches 95; Conservative 82; Mismatches 141; Indels 211; Gaps 27;

Qy 4 LENRVLGSLVLMISGL-----AATAPVAEPSPGYTLERVILSRHGVRSSTKQT 54

Db 320 IENQLRSEGVMLTDLILKHAKEAFQATVSTADTGLKEA---MEKV-----QKYN 365

Qy 55 QLANDVTTPKWPQKVPKAGVLTTPRGAEVLTMGFGYGVDFRSGLLAAGCP----- 105

Db 366 QLMRDPPLDEL-----LSATLT-----KVQESIQIFGLHNLKRLI-----CPYFIRALPL 413

Qy 106 --AEGGYAQAIDQRTLTGQAFDGVAPGCGLTVHNOAD-----LKTDPPLPHVETGV 159

Db 414 VEAISG-----DLDE-----VLHRLLPGLTELKLDYEEFKGVKMQAGSIFRAWDESI 460

Qy 160 CKLDNA-----QTDKATEERLGGPLDTSQRYAKPPAQMGDVLPFAASPYCKSLQQQG 212

Db 461 KEFTNVAEVRTRRNEKFI-----PI-KINPHASLQSLDYVHFRDN----- 503

Qy 213 KTCDFAHFAANEVNVNKEGKVTLSGLSPLSSLTGIFL--LQNAQAMPEV--AWQRLK- 267

Db 504 -----HEQRTIINVLGPKATVNGIVTASGANGVAVVEEIGDVADEVKQAEALKD 557

Qy 268 -----CAENWVSLSLHN-----AQFILMA 287

Db 558 VLLDCTRETEKRWRAENIYNERTARVENSIIARLDRLATKANENMFRVFSKFNALF 617

Qy 288 KTP-----YIARHGTPLLQOITDALTQLDAQCKLPISAQNRVFLFGH----- 333

Db 618 VRPKIRGAIAYQ-TQLIDNVKQAIS-----SLHERFKQYGHSEAHAWAQLH 664

Qy 334 -----DTNNTANTAGMLGADW-----QLPEQPD-----NTPPGG 361

Db 665 DLPPVSGAIIWARQIBERQLDQYMKVEQVLGSDWALHTEGQKLQNESDLFRKCLDTRP-- 722

Qy 362 GLYFELWQNPDNHQRVAVKMFYQTMOLNVAEKLDLKNPAGIISVAV 410

Db 723 --IFEAWLH-DVORKQISISGLLFTINRISA-----GNILELAV 759

RESULT 14

VIE2 HCMVA STANDARD; PRT; 411 AA.

AC P1983;

DT 01-FEB-1991 (Rel. 17, Created)

DT 01-FEB-1991 (Rel. 17, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE 45 kDa immediate-early protein 2 (IE2) (U1122 protein).

GN U1122.

OS Human cytomegalovirus (strain AD169).

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

OC Betaherpesvirinae; Cytomegalovirus.

OX NCBI_TaxID=10360;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=90269039; PubMed=2161319;

Chae M.S., Bankier A.T., Beck S., Bohni R., Brown C.M., Cerny R.,
Horsnell T., Hutchison S.A. III, Kouravides T., Martignetti J.A.,
Freddie E., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.;
"Analysis of the protein-coding content of the sequence of human
cytomegalovirus strain AD169,"
Curr. Top. Microbiol. Immunol. 154:125-169 (1990).
-!- FUNCTION: ACTIVATES THE IE1-7 PROMOTER. THIS ACTIVATION IS
AUGMENTED BY THE IE1 PROTEIN. IT DOWN-REGULATES THE TRANSCRIPTION
OF GENES UNDER THE CONTROL OF THE MAJOR IE PROMOTER.
-!- SIMILARITY: OF ITS C-TERMINUS TO THE C-TERMINUS OF THE MURINE
CYTOMEGALOVIRUS IMMEDIATE-EARLY PROTEIN IE3.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
or send an email to license@isb-sib.ch.

EMBL; X17403; CAA35324.1; -.
PIR; S09889; EDDEM4.
InterPro; IPR005028; Herpes IE2_3.
Pfam; PF03361; Herpes IE2_3_1_
Early protein; Transcription regulation; Activator; Zinc-finger;
DNA-binding. 89 151 GLY/SBR-RICH.
DOMAIN ZN_FING 258 284 POTENTIAL.
SEQUENCE 411 AA; 44831 MW; 14415DA5AD2C7F31 CRC64;

Query Match 4.5%; Score 98.5; DB 1; Length 411;
Best Local Similarity 25.3%; Pred. No. 2,7;
Matches 78; Conservative 30; Mismatches 119; Indels 81; Gaps 19;

QY 87 GGFGDYFRSLGLLAACPAEAGVGAQAIDORTLTGQAFLDGVAPGCCGLTVHQADLK 146
Ddb ||||| :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
128 GGFGG--AASSLSLLCGHSGSSGASTGPKKKSKRI-----SELDNE 167
QY ||||| :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
147 KTDFLFHFVETGVCKLDNAQTDXAIERLGGPLDTVSQRYAK-----PFA--QM 194
Ddb ||||| :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
168 KVRNIMDKXTFFC-TFNVTQRGRVK-----IDSVSRMFRNTSRLEYKNLPPTIPSMH 221
QY 195 DYLNFAASPCKSLOOGKTCDPAFAANEVNNVNKETKVLSG--PLASSLTGEIFLL 252
Ddb ||||| :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
222 QVLDEATKA-CKTMQVNNKIQTIIYTRNHEVKSVDVAVRCRLGTWCNLALSTP----FLM 276
QY ::::: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
253 NQAQAM--PEVAWQLKGAE-----WVSLSLSLHNAQF-----NLMAKTPYIARH 295
Ddb :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
277 EHTFPVTHPEVA-QRTADACNEGVAAM-SLKLHTHLQCPRSSDYRNMI-----H 327
QY 296 KGTP--LIQIDITALTILDCAQCKLPISAQNEVFLF--GGHDFTIANIAGMLGADWOLF 351
Ddb ||||| :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
328 AATPVDDLGLNALCLPLM-----QKPFKQVNRIFSTNQGGFMLPIYETAKAYAVQG?- 381
QY 352 EQPDNTPP 359
Ddb ||||| :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
382 EQPTETPP 389

RESULT 15.
CYDD_HAEIN STANDARD; PRT; 586 AA.
AC P45082;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Transport ATP-binding protein cydd.
CYDD OR H1157
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
NCBI_TaxID=727;
RX [1]

QY 184 QRYAKPFAOMGDVLPALSPYCKSLQOQKTCDFAPFAANVNVNKEGTVTLSPPLALS 243
Db 384 QSGAGKTSIMNVILGFL--PYEGSLKINGQBELRESNLADWRKHIWVG-----QNPLLLQ 436
QY 244 STLGEIFLLQNAQAMPEVAVORL--KGAENWVSLSLHN-----AQFNLMARKTP 290
Db 437 GTIKENLLGDVQANDEEINQALMESQAKEFTDKLGLHHEIKDGLGISVGQAORLA--- 493
QY 291 YIAR---HKGTPLLOQIDTALTQLDAQOKLIPISAQNRVFLGGHDTNI 337
Db 494 -IARALLRKGDILLLEDEPTA---SLDAQSENVLQALNEA---SQHQTTL 536

Search completed: April 30, 2004, 12:39:27
Job time : 13.2439 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 30, 2004, 12:34:35 ; Search time 38.0488 Seconds
(without alignments)
3482.836 Million cell updates/sec

Title: US-10-021-723B-4
Perfect score: 2200
Sequence: 1 MSVLENRVRLSGLVLMISGL.....NPAGIISVAVACENNGDDK 420

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL 25.*
- 1: sp_archaea.*
 - 2: sp_bacteria.*
 - 3: sp_fungi.*
 - 4: sp_human.*
 - 5: sp_invertebrate.*
 - 6: sp_mammal.*
 - 7: sp_mhc.*
 - 8: sp_organalle.*
 - 9: sp_phage.*
 - 10: sp_plant.*
 - 11: sp_rodent.*
 - 12: sp_virus.*
 - 13: sp_vertebrate.*
 - 14: sp_unclassified.*
 - 15: sp_virus.*
 - 16: sp_bacteriap.*
 - 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2183	99.2	441	16	Q8ZFP6
2	941.5	42.8	432	2	Q8RKD7
3	935.5	42.5	432	2	Q8GN88
4	933.5	42.4	432	2	Q8RKD6
5	933.5	42.4	432	2	Q8RKD8
6	933.5	42.4	432	2	Q8RXE0
7	932.5	42.4	432	2	Q8RKD9
8	929.5	42.2	432	2	Q8RKD5
9	926.5	42.1	434	16	Q8XC29
10	926.5	42.1	446	16	Q8CW75
11	919.5	41.8	432	16	Q7UD08
12	919.5	41.8	442	16	Q83RW2
13	532	24.2	413	16	Q8XBZ6
14	527	24.0	413	16	Q7UD02
15	527	24.0	421	16	Q83RV6
16	514.5	23.4	392	16	Q8PP76

17	504	22.9	414	16	Q9AAQ4
18	500	22.7	443	16	Q8P330
19	495	22.5	413	16	Q8Z7P1
20	432	19.6	428	2	Q8GD20
21	429.5	19.5	421	2	Q84CN9
22	422	19.2	318	16	Q8CW72
23	421.5	19.2	421	2	Q7WSY1
24	413	18.8	435	16	Q8PF53
25	332	15.1	414	16	Q8PPX6
26	206.5	9.4	419	2	Q8VQS2
27	156	7.1	435	5	Q8MY06
28	127.5	5.8	531	5	Q8GNZ3
29	125.5	5.7	447	5	Q9TW17
30	124.5	5.7	447	5	Q9USU4
31	124.5	5.7	447	5	Q9USU1
32	124.5	5.7	447	5	Q9USV0
33	123.5	5.6	447	5	Q9USU2
34	123.5	5.6	447	5	Q9USU8
35	123.5	5.6	447	5	Q9US58
36	123.5	5.6	447	5	Q9US70
37	123.5	5.6	447	5	Q9IVM1
38	123.5	5.6	447	5	Q9US73
39	123.5	5.6	447	5	Q9TVX2
40	123	5.6	447	5	Q9US75
41	122.5	5.6	447	5	Q9USU5
42	122.5	5.6	447	5	Q97186
43	122.5	5.6	447	5	Q9USU3
44	122.5	5.6	447	5	Q9TW53
45	122.5	5.6	447	5	Q97187

ALIGNMENTS

RESULT 1

Q8ZFP6	PRELIMINARY;	PRT;	441	AA.
AC	Q8ZFP6;			
DT	01-MAR-2002 (T-EMBLrel. 20, Created)			
DT	01-MAR-2002 (T-EMBLrel. 20, Last sequence update)			
DT	01-JUN-2003 (T-EMBLrel. 24, Last annotation update)			
DE	Probable histidine acid phosphatase (EC 3.1.3.2) (Phosphoanhydride phosphorylase).			
GN	YPO1648 OR APPA OR Y1810.			
OS	Yersinia pestis.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Yersinia.			
OX	NCBI_TaxID=632;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CO-92 / Biovar Orientalis;			
RX	MEDLINE=21470413; PubMed=11586360;			
RA	Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G., Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L., Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M., Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G., Feilwell T., Hamlin N., Holroyd S., Jagers K., Karlyshev A.V., Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.; "Genome sequence of Yersinia pestis, the causative agent of plague."			
RL	Nature 413:523-527(2001).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=KIMS / Biovar Mediaevalis;			
RX	MEDLINE=22137863; PubMed=12142430;			
RA	Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P., Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C., Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V., Straley S.C., McDonough K.A., Niles M.L., Matson J.S., Blattner F.R., Perry R.D.; "Genome sequence of Yersinia pestis KIM."			
RL	J. Bacteriol. 184:4601-4611(2002).			
DR	EMBL; AJ414149; CAC90470.1; -.			

Q9aaq4	caulobacter
Q9p330	xanthomonas
Q8z7p1	salmonella
Q8gd20	pseudomonas
Q84cn9	klebsiella
Q8cw72	escherichia
Q7wsy1	klebsiella
Q8pf53	xanthomonas
Q8ppx6	xanthomonas
Q8vqs2	klebsiella
Q8my06	drosophila
Q8gnz3	leishmania
Q9tw17	drosophila
Q9us04	drosophila
Q9us01	drosophila
Q9usv0	drosophila
Q9us02	drosophila
Q9us08	drosophila
Q9us58	drosophila
Q9us70	drosophila
Q9ivm1	drosophila
Q9us73	drosophila
Q9tvx2	drosophila
Q9us75	drosophila
Q9us05	drosophila
Q97186	drosophila
Q9us03	drosophila
Q9tw53	drosophila
Q97187	drosophila

RESULT 3	Q8GN88	PRELIMINARY;	PRT;	432 AA.
Q8GN88	AC	Q8GN88;		
DT	01-MAR-2003	(TrEMBLrel. 23, Created)		
DT	01-MAR-2003	(TrEMBLrel. 23, last sequence update)		
DT	01-OCT-2003	(TrEMBLrel. 25, last annotation update)		
DE	APPA.			
GN	GN	APPA.		
OS	Escherichia coli;			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Escherichia.			
OX	NCBI_TaxID=562;			
[1]	RN	SEQUENCE FROM N.A.		
RP	Zhang Z., He J., Yao B., Zhou Y., Chen Y., Yi Y.;			
RA	"Production of phycase and acid phosphatase by use of silkworm-			
RT	bio reactor.";			
RL	Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Chen Y., Zhu Z., Zhang Z., He J.;			
RT	"Cloning and overexpression of phycase gene appA from Escherichia			
RT	coli.";			
RL	Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.			
RL	EMBL; AF537219; AAN28334.1; -			
DR	GO; GO:0003933; F:acid phosphatase activity; IEA.			
DR	InterPro; IPR000560; HisAc_phsphtase.			
DR	Pfam; PF00328; acid phosphat; 1.			
DR	PROSITE; PS00616; HIS ACID PHOSPHAT 1; 1.			
DR	PROSITE; PS00778; HIS ACID PHOSPHAT 2; 1.			

SQ SEQUENCE 432 AA; 47055 MW; 5B355D76E7377737 CRC64;

Query Match 42.5%; Score 935.5; DB 2; Length 432;
Best Local Similarity 47.7%; Pred. No. 1.4e-67;
Matches 195; Conservative 62; Mismatches 137; Indels 15; Gaps 7;

QY 13 LVLMGLAATAPVA---ABPSGYTLERVVLSRHGVRSPKOTQLMNDVTPDKWPWP 69
DB 5 LIPFLSLLIPLTPQSAFAQSEPE-LKLESVVIVSRHGVRAPTKATQLMQDVTDPDWTWP 63
QY 70 VKAGVILPRGAEVLTMGFGYDYSRISGLIA-ACPAEGGVYAQADIDORTLITGOAFL 128
DB 64 VKLGWLTFRGGELIAYLGHYQRLVADGLLAKGCPQSGQVAILADVDEATRTKGEAFA 123
QY 129 DGVAFCGLTVHNOADLKKTKDPLFHPVETGVCKLDNAQDKAIEERLGGPLDVTVSRYAK 188
DB 124 AGLAPDCAITVHTQADTSSPDPLFNPLKTVGCQLDNNANVTDAILSRAGGSIAFTGHROT 183
QY 189 PFAQMGDVINFAASPYCKSLQOQKTCDFAPHAFAENVNKNKGTVTLSGPIALSSTLGE 248
DB 184 AFRELERVLNFPQSNLCKREKQDESCSLTQALPSELKVSAD--NVSLTGAVSLASMLTK 241
QY 249 IFLQNAQAMPEVAWORLKGAEVWVLSLSLHNAQFNLMKTFYIARHKGTPLLQOQIDTAL 308
DB 242 IFLQQAQGMPEPGWGRITDHSQWNTLLSLHNAQFYLLQRTPEVARGRATPLDLIKTAL 301
QY 309 T---LQIDAQOGKLPISAQNRVFLGQHDNTIANIAGMLGADWOLPEQDNTPPGGGLVP 365
DB 302 THPPPKQAYGVTLPTS-----VLFAGHDTNLANLGGALELNTLPGQPDNTPPGGELVF 357
QY 366 ELWQNDPNHQRVAVKMFYQTMQDLNRAEKLDLKNPAGIISVAVAGCE 414
DB 358 ERWRRLSDNSQWIVSLVFQTLQQRDKTPLSL-NTPPGEVKLTLAGCE 405

RESULT 4

Q8RKD6 PRELIMINARY; PRT; 432 AA.

ID Q8RKD6
AC Q8RKD6;
DT 01-JUN-2002 (TremBLrel. 21, Created)
DT 01-JUN-2002 (TremBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)
DE Periplasmic phosphoanhydride phosphohydrolase.
GN APPA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90368616; PubMed=2168385;
RA Dassa J., Marck C., Boquet P.-L.L.;
RT "The complete nucleotide sequence of the Escherichia coli gene appa
RT reveals significant homology between pH 2.5 acid phosphatase and
RT glucose-1-phosphatase.";
RL J. Bacteriol. 172:5497-5500 (1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93054596; PubMed=1429631;
RA Ostanin K., Harms E.H., Stevis P.E., Zhou M.M.,
RA Van Etten R.L.;
RT "Overexpression, site-directed mutagenesis, and mechanism of
RT Escherichia coli acid phosphatase.";
RL J. Biol. Chem. 267:22830-22836 (1992).
DR EMBL; L03374; AAA00006.1;
DR GO; GO:0003993; F:acid phosphatase activity; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR000560; Hisac_phsphtse.
DR Pfam; PF00328; acid_phosphat; 1.
DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_1; 1.
DR Hydrolase.
KW Hydrolase.
SQ SEQUENCE 432 AA; 46971 MW; AF86C41EA6193AC5 CRC64;

Query Match 42.4%; Score 933.5; DB 2; Length 432;
Best Local Similarity 47.7%; Pred. No. 2e-67;
Matches 195; Conservative 61; Mismatches 138; Indels 15; Gaps 7;

QY 13 LVLMGLAATAPVA---ABPSGYTLERVVLSRHGVRSPKOTQLMNDVTPDKWPWP 69
DB 5 LIPFLSLLIPLTPQSAFAQSEPE-LKLESVVIVSRHGVRAPTKATQLMQDVTDPDWTWP 63
QY 70 VKAGVILPRGAEVLTMGFGYDYSRISGLIA-ACPAEGGVYAQADIDORTLITGOAFL 128
DB 64 VKLGWLTFRGGELIAYLGHYQRLVADGLLAKGCPQSGQVAILADVDEATRTKGEAFA 123
QY 129 DGVAFCGLTVHNOADLKKTKDPLFHPVETGVCKLDNAQDKAIEERLGGPLDVTVSRYAK 188
DB 124 AGLAPDCAITVHTQADTSSPDPLFNPLKTVGCQLDNNANVTDAILSRAGGSIAFTGHROT 183
QY 189 PFAQMGDVINFAASPYCKSLQOQKTCDFAPHAFAENVNKNKGTVTLSGPIALSSTLGE 248
DB 184 AFRELERVLNFPQSNLCKREKQDESCSLTQALPSELKVSAD--NVSLTGAVSLASMLTE 241
QY 249 IFLQNAQAMPEVAWORLKGAEVWVLSLSLHNAQFNLMKTFYIARHKGTPLLQOQIDTAL 308
DB 242 IFLQQAQGMPEPGWGRITDHSQWNTLLSLHNAQFYLLQRTPEVARGRATPLDLIKTAL 301
QY 309 T---LQIDAQOGKLPISAQNRVFLGQHDNTIANIAGMLGADWOLPEQDNTPPGGGLVP 365
DB 302 THPPPKQAYGVTLPTS-----VLFAGHDTNLANLGGALELNTLPGQPDNTPPGGELVF 357
QY 366 ELWQNDPNHQRVAVKMFYQTMQDLNRAEKLDLKNPAGIISVAVAGCE 414
DB 358 ERWRRLSDNSQWIVSLVFQTLQQRDKTPLSL-NTPPGEVKLTLAGCE 405

RESULT 5

Q8RKD8 PRELIMINARY; PRT; 432 AA.

ID Q8RKD8
AC Q8RKD8;
DT 01-JUN-2002 (TremBLrel. 21, Created)
DT 01-JUN-2002 (TremBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)
DE Periplasmic phosphoanhydride phosphohydrolase.
GN APPA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90368616; PubMed=2168385;
RA Dassa J., Marck C., Boquet P.-L.L.;
RT "The complete nucleotide sequence of the Escherichia coli gene appa
RT reveals significant homology between pH 2.5 acid phosphatase and
RT glucose-1-phosphatase.";
RL J. Bacteriol. 172:5497-5500 (1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93054596; PubMed=1429631;
RA Ostanin K., Harms E.H., Stevis P.E., Zhou M.M.,
RA Van Etten R.L.;
RT "Overexpression, site-directed mutagenesis, and mechanism of
RT Escherichia coli acid phosphatase.";
RL J. Biol. Chem. 267:22830-22836 (1992).
DR EMBL; L03372; AAA00004.1;
DR GO; GO:0003993; F:acid phosphatase activity; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR000560; Hisac_phsphtse.
DR Pfam; PF00328; acid_phosphat; 1.
DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
DR Hydrolase.
KW Hydrolase.
SQ SEQUENCE 432 AA; 46971 MW; 9A85536B57FCCFB5 CRC64;

Best Local Similarity 47.7%; Pred. No. 2e-67;
Matches 195; Conservative 61; Mismatches 138; Indels 15; Gaps 7;

QY 13 LVLMISGLAIAITAPVA---AEPGTYTLERVVILSRHGVRSPKTKQTOLMNDVTPDKWPQP 69
DB 5 LIPFLSLLIPLTPQSAFAQSEPE-LKLESVVIVSAHGVRAPTKATQLMQDVTDPDWTWP 63
QY 70 VKAGYTLPRGAELVTLMGGFYGYFRSLGLLA-AGCPAEGVVAQADIDORTLRTGQAF 128
DB 64 VKLGWLTTPRGELIAYLGHYQORLQVADGLLAKKGCPCQGVVAIIADVDERTRKTGEAFA 123
QY 129 DGVAPEGGLTVHNOADLKKTDPLFHPVETGVCKLDNAQTDKAIERLGGPLDTVSQRYAK 188
DB 124 AGLAPDCAITVHTQADTSSPDPLFNLKTVGCOLDNANVTDAILSRAGSIADFTGHRQT 183
QY 189 PFAQMGDVLPFAASPYCKSLQOQKTCDFAHFAANVNNKGTGYTLSPGLALSTLGE 248
DB 184 AFRELERVLNFPQSNLCKREKQDECSLTQALPSELKVSAD--NVSLTAGVSLASMLTE 241
QY 249 IFLQNAQAMPEVANQELKGAENWVSLSLHNAQFNLMAKTPYIARHKGTPLLQOQIDTAL 308
DB 242 IFLQQAQMGPEPGWGRITDHSQWNTLLSLHNAQFYLLQRTPEVARSTRATPLDLIKTAL 301
QY 309 T---LQIDAQOQKLPISAQNRVFLFGGHDNTNIAAGMLGADWQLPEQPDNTPPGGGLVF 365
DB 302 TPHPPOKQAGVTLPTS---VLFIAGHDTNLANLGGALELNLWTLPGQPDNTPPGGGLVF 357
QY 366 ELWQNPNDHORYVAVKMFYQTMOLRNAEKLDKNNPAGIISVAVAGCE 414
DB 358 ERWRRLSDNSQWIOVSLVFQTLQOQMRDKTPLSL-NTPPGEVKLTLAGCE 405

RESULT 6
Q8RKD9 PRELIMINARY; PRT; 432 AA.

ID Q8RKD9; Q8RKD9; PRT; 432 AA.
AC Q8RKD9; Q8RKD9; PRT; 432 AA.
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Periplasmic phosphoanhydride phosphohydrolase.
GN APPA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90368616; PubMed=2168385;
RA Daasa J., Marck C., Boquet P.-L.L.;
RT "The complete nucleotide sequence of the Escherichia coli gene appA
RT reveals significant homology between pH 2.5 acid phosphatase and
RT Glucose-1-phosphatase."
RL J. Bacteriol. 172:5497-5500(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93054596; PubMed=1429631;
RA Ostain K., Harms E.H., Stevis P.E., Kuciel R., Zhou M.M.,
RA Van Etten R.L.;
RT "Overexpression, site-directed mutagenesis, and mechanism of
RT Escherichia coli acid phosphatase."
RL J. Biol. Chem. 267:22830-22836(1992).
DR EMBL; L03371; AAA00002.1;
DR GO; GO:0003993; F:acid phosphatase activity; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR000560; HisAc_phsphtse.
DR Pfam; PF00328; acid_phosphat; 1.
DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
KW Hydrolase.
SQ SEQUENCE 432 AA; 46971 MW; 5BB0632D4682EABF CRC64;

Query Match 42.4%; Score 933.5; DB 2; Length 432;
Best Local Similarity 47.7%; Pred. No. 2e-67;
Matches 195; Conservative 61; Mismatches 138; Indels 15; Gaps 7;

QY 13 LVLMISGLAIAITAPVA---AEPGTYTLERVVILSRHGVRSPKTKQTOLMNDVTPDKWPQP 69
DB 5 LIPFLSLLIPLTPQSAFAQSEPE-LKLESVVIVSAHGVRAPTKATQLMQDVTDPDWTWP 63
QY 70 VKAGYTLPRGAELVTLMGGFYGYFRSLGLLA-AGCPAEGVVAQADIDORTLRTGQAF 128
DB 64 VKLGWLTTPRGELIAYLGHYQORLQVADGLLAKKGCPCQGVVAIIADVDERTRKTGEAFA 123
QY 129 DGVAPEGGLTVHNOADLKKTDPLFHPVETGVCKLDNAQTDKAIERLGGPLDTVSQRYAK 188
DB 124 AGLAPDCAITVHTQADTSSPDPLFNLKTVGCOLDNANVTDAILSRAGSIADFTGHRQT 183
QY 189 PFAQMGDVLPFAASPYCKSLQOQKTCDFAHFAANVNNKGTGYTLSPGLALSTLGE 248
DB 184 AFRELERVLNFPQSNLCKREKQDECSLTQALPSELKVSAD--NVSLTAGVSLASMLTE 241
QY 249 IFLQNAQAMPEVANQELKGAENWVSLSLHNAQFNLMAKTPYIARHKGTPLLQOQIDTAL 308
DB 242 IFLQQAQMGPEPGWGRITDHSQWNTLLSLHNAQFYLLQRTPEVARSTRATPLDLIKTAL 301
QY 309 T---LQIDAQOQKLPISAQNRVFLFGGHDNTNIAAGMLGADWQLPEQPDNTPPGGGLVF 365
DB 302 TPHPPOKQAGVTLPTS---VLFIAGHDTNLANLGGALELNLWTLPGQPDNTPPGGGLVF 357
QY 366 ELWQNPNDHORYVAVKMFYQTMOLRNAEKLDKNNPAGIISVAVAGCE 414
DB 358 ERWRRLSDNSQWIOVSLVFQTLQOQMRDKTPLSL-NTPPGEVKLTLAGCE 405

RESULT 7
Q8RKD9 PRELIMINARY; PRT; 432 AA.

ID Q8RKD9; Q8RKD9; PRT; 432 AA.
AC Q8RKD9; Q8RKD9; PRT; 432 AA.
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Periplasmic phosphoanhydride phosphohydrolase.
GN APPA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90368616; PubMed=2168385;
RA Daasa J., Marck C., Boquet P.-L.L.;
RT "The complete nucleotide sequence of the Escherichia coli gene appA
RT reveals significant homology between pH 2.5 acid phosphatase and
RT Glucose-1-phosphatase."
RL J. Bacteriol. 172:5497-5500(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93054596; PubMed=1429631;
RA Ostain K., Harms E.H., Stevis P.E., Kuciel R., Zhou M.M.,
RA Van Etten R.L.;
RT "Overexpression, site-directed mutagenesis, and mechanism of
RT Escherichia coli acid phosphatase."
RL J. Biol. Chem. 267:22830-22836(1992).
DR EMBL; L03371; AAA00003.1;
DR GO; GO:0003993; F:acid phosphatase activity; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR000560; HisAc_phsphtse.
DR Pfam; PF00328; acid_phosphat; 1.
DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
KW Hydrolase.
SQ SEQUENCE 432 AA; 47033 MW; 9F29B9DF9C368175 CRC64;

Query Match 42.4%; Score 932.5; DB 2; Length 432;
Best Local Similarity 47.7%; Pred. No. 2.4e-67;
Matches 195; Conservative 62; Mismatches 137; Indels 15; Gaps 7;

QY 13 LVLMISGLAIAITAPVA---AEPGTYTLERVVILSRHGVRSPKTKQTOLMNDVTPDKWPQP 69

Db 5 LIPFLSLPLTPQSAFQSEPE-LKLESVIVSRNGVRAFTKATQLMQDVTDPDWTWP 63
 QY 70 VKAGVLTTPRGAEVLTLMGGFYGDYFRSLGLLA-AGCPAEGGVYAQADIDQRTLTGQAF 128
 Db 64 VKLGLWLTTPRGELIAYLGHYQRLVADGLLAKKGCPSQGVALLIADVDERTKTGEAFA 123
 QY 129 DGVAFCGLTVHNOADLKKTDPLRHPVETGVCKLNAQTDKAIERLGGPLDVTVSQRYAK 188
 Db 124 AGLAPDCAITVHTQADTSSPDPLFNPLKTVGCQLDNANVTDAILSRAGGSIAFTGHRQT 183
 QY 189 PPAQMGDVINFAASPYCKSLQOQKTCDFAHFAANVNVNKEGTVKVLGSLALSSTLGE 248
 Db 184 AFRELERVLPQSNLCIKREKQDESCSLTQALPSELKVSAD--NVSLTGAVSLASMLTE 241
 QY 249 IFLQNAQAMPEVAVQRLKGAENWVLSLSLHNAQNLMAKTPYIARHKGTPLLQOQIDTAL 308
 Db 242 IFLQNAQAMPEVAVQRLKGAENWVLSLSLHNAQNLMAKTPYIARHKGTPLLQOQIDTAL 301
 QY 309 T---LQDQAQOKLPIAQNVRVFLFGGHDNTNANTAGMLGADWQLPQPDNTPPGGGLVF 365
 Db 302 TPHPKQKAYGVTLPTS----VLFAGADTNLANTAGMLGADWQLPQPDNTPPGGGLVF 357
 QY 366 ELWQNPNDHQRVAVKMFYQTMQDLRQNAEKLDKNNPAGIISVAVAGCE 414
 Db 358 ERWRRLSDNSQWVLSVFTLQQRDKTPLSL-NTPPGEVKLTLAGCE 405

RESULT 8

Q8XKD5 PRELIMINARY; PRT; 432 AA.
 ID Q8XKD5
 AC Q8XKD5
 DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
 DE Periplasmic phosphoanhydride phosphohydrolase.
 GN APPA.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90368616; PubMed=2168385;
 RA Dassa J., March C., Boquet P.-L.L.;
 RT "The complete nucleotide sequence of the Escherichia coli gene appA
 RT reveals significant homology between pH 2.5 acid phosphatase and
 RT Glucose-1-phosphatase.";
 RL J. Bacteriol. 172:5497-5500 (1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93054596; PubMed=1429631;
 RA Ostanin K., Harms E.H., Stevis P.E., Kuciel R., Zhou M.M.,
 RA Van Etten R.L.;
 RT "Overexpression, site-directed mutagenesis, and mechanism of
 RT Escherichia coli acid phosphatase.";
 RL J. Biol. Chem. 267:22830-22836 (1992).
 DR EMBL; L03375; AAA00007.1;
 DR GO; GO:0003993; F:acid phosphatase activity; IEA.
 DR GO; GO:0016787; F:hydrolase activity; IEA.
 DR InterPro; IPR000560; HisAc_phsphtse.
 DR Pfam; PF00328; acid_phosphat; 1.
 DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
 KW Hydrolase.
 SQ SEQUENCE 432 AA; 46990 MW; 951P393EA9A1A47C CRC64;

Query Match 42.2%; Score 929.5; DB 2; Length 432;
 Best Local Similarity 47.7%; Pred. No. 4.2e-67;
 Matches 195; Conservative 61; Mismatches 138; Indels 15; Gaps 7;
 QY 13 LVLMLSLAATAPVA---AEPGTYLRLVWILSRHGVSPKQTLQNDVTPDKWPWP 69
 Db 5 LIPFLSLPLTPQSAFQSEPE-LKLESVIVSRNGVRAFTKATQLMQDVTDPDWTWP 63

QY 70 VKAGVLTTPRGAEVLTLMGGFYGDYFRSLGLLA-AGCPAEGGVYAQADIDQRTLTGQAF 128
 Db 64 VKLGLWLTTPRGELIAYLGHYQRLVADGLLAKKGCPSQGVALLIADVDERTKTGEAFA 123
 QY 129 DGVAFCGLTVHNOADLKKTDPLRHPVETGVCKLNAQTDKAIERLGGPLDVTVSQRYAK 188
 Db 124 AGLAPDCAITVHTQADTSSPDPLFNPLKTVGCQLDNANVTDAILSRAGGSIAFTGHRQT 183
 QY 189 PPAQMGDVINFAASPYCKSLQOQKTCDFAHFAANVNVNKEGTVKVLGSLALSSTLGE 248
 Db 184 AFRELERVLPQSNLCIKREKQDESCSLTQALPSELKVSAD--NVSLTGAVSLASMLTE 241
 QY 249 IFLQNAQAMPEVAVQRLKGAENWVLSLSLHNAQNLMAKTPYIARHKGTPLLQOQIDTAL 308
 Db 242 IFLQNAQAMPEVAVQRLKGAENWVLSLSLHNAQNLMAKTPYIARHKGTPLLQOQIDTAL 301
 QY 309 T---LQDQAQOKLPIAQNVRVFLFGGHDNTNANTAGMLGADWQLPQPDNTPPGGGLVF 365
 Db 302 TPHPKQKAYGVTLPTS----VLFAGADTNLANTAGMLGADWQLPQPDNTPPGGGLVF 357
 QY 366 ELWQNPNDHQRVAVKMFYQTMQDLRQNAEKLDKNNPAGIISVAVAGCE 414
 Db 358 ERWRRLSDNSQWVLSVFTLQQRDKTPLSL-NTPPGEVKLTLAGCE 405

RESULT 9

Q8XC29 PRELIMINARY; PRT; 434 AA.
 ID Q8XC29
 AC Q8XC29
 DT 01-MAR-2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Phosphoanhydride phosphorylase, pH 2.5 acid phosphatase,
 DE periplasmic.
 GN APPA OR Z1397 OR ECS1136.
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
 RL Nature 409:529-533 (2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22 (2001).
 DR EMBL; AB005292; AAG55528.1; ALT_INIT.
 DR PIR; AP002554; BAB34559.1; -;
 DR PIR; D85633; D85633.
 DR PIR; H90770; H90770.
 DR GO; GO:0003993; F:acid phosphatase activity; IEA.
 DR InterPro; IPR000560; HisAc_phsphtse.
 DR Pfam; PF00328; acid_phosphat; 1.
 DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
 DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
 KW Complete proteome.

SEQ	SEQUENCE	434 AA; 47337 MW; F197DFTD718659P9C4	CRG64;
Query Match	42.1%;	Score 926.5; DB 16; Length 434;	
Best Local Similarity	47.3%;	Pred. No. 7.4e-67;	
Matches 194;	Conservative	62; Mismatches 139; Indels 15; Gaps	
QY	13	LVLMGLSLGAIAATPAV--AEPSS-GYTLERVVILSHGVRSPTKOTQLMNDVTPDKYQW	68
DB	5	LIPFLSLLIPITPOSAFAQSEPEPELKLSSVIVSRHGVRAFTKATQMDQVTPDAMPNW	64
QY	69	PVKAGYLTTPRGAEVLVTLMGFGYGDYFRSLGLLA-AGCPAEGGVYAQADIDORTLTQOAF	127
DB	65	PVKLGWLTTPRGELIAYLGHYQRLVADGLLTKKGCPQPGQVIAIDVDERTKTKGEAF	124
QY	128	LDGVAPGCGLVNHNQADLKTDPLEFPVETGVCKLDNAQTKDAIEERLGGPLDVTQSYA	187
DB	125	AAGLAPDCAIVHTQARTSPDFLNFELKTVGCQLDNNVTDAILSRAGGSIADF7GHRQ	184
QY	188	KPFQMGQVLMNFAASPYCKSLQOQKTKCFAPHAPEANVNVNKGTVKTLSPGLALSSTLG	247
DB	185	TAFRELERVLNFPQSNCLNREKQDESCLTQALPSELKVSAD--NVSLTGAVSLASMLT	242
QY	248	EIFLLQNAQAMPEVAWORLKGAEWVSLLSLHNAQNLMAKTPVIARHKGTPLLQQIDTA	307
DB	243	EIFLLQQAQNPFGWGRITDSHQWVTLSLHNAQYLLSQRTPEVARGRAIPLHDLNIA	302
QY	308	LT---LQLDAGQCKLPTISAQNRVFLFGCHDTNFIANAGMLGADWOLPEQPDNTPPGGGLV	364
DB	303	LTPHPQQAQYGVTLPTS---VLFIAGHDTNLNGLGALNLNWTLPQPDNTPPGGELV	358
QY	365	FELQNPNDHORYVAVAKMFTYQTMQLRKAELKDLKNPAGIISVAVACE	414
DB	359	FERWRSLSDNSQWTVQSLVFTQLQMRDKTPLSL-NTPPGEVKTIACGE	407
RESULT 10			
Q8CW75		PRELIMINARY; PRT; 446 AA.	
ID	Q8CW75		
AC	Q8CW75		
DT	01-NAR-2003 (TrenBrel. 23, Created)		
DT	01-NAR-2003 (TrenBrel. 23, Last sequence update)		
DT	01-OCT-2003 (TrenBrel. 25, Last annotation update)		
DE	Periplasmic appA protein precursor.		
GN	APPA OR C121.		
OS	Escherichia coli O6.		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
OC	Enterobacteriaceae; Escherichia.		
OX	NCBI_TaxID=217992;		
RN	[1]_TaxID=217992;		
RC	SEQUENCE FROM N.A.		
RP	STRAIN=O6:HI / CFT073 / ATCC 700928;		
RX	MEDLINE=22388234; PubMed=12471157;		
RA	Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,		
RA	Rasko D., Buckles E.J., Liou S.-R., Boutin A., Hackett J., Stroud D.,		
RA	Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,		
RA	Mobley H.L.T., Donnenberg M.S., Blattner F.R.;		
RT	"Extensive mosaic structure revealed by the complete genome sequence		
RT	of uropathogenic Escherichia coli."		
RL	Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).		
RL	EMBL; AF016758; AAN79589.1; -.		
DR	GO; GO:0003993; F:acid phosphatase activity; IEA.		
DR	InterPro; IPR000560; HisAc phosphatase.		
DR	Pfam; PF00328; acid_phosphat_1		
DR	PROSITE; PS00616; HIS-ACID_PHOSPHAT_1; 1.		
DR	PROSITE; PS00778; HIS-ACID_PHOSPHAT_2; 1.		
KW	Complete proteome.		
SEQ	SEQUENCE 446 AA; 48605 MW; F1308CDG91DB9F49	CRG64;	

Query Match 42.1%; Score 926.5; DB 16; Length 446;
Best Local Similarity 47.2%; Pred. No. 7.7e-87;
Matches 193; Conservative 63; Mismatches 138; Indels 15; Gaps 7;
QY 13 LVLMKSLGAATAPVA---AEPGYTLERVVLSRHGVSPKTKQTQLANDVTPDKWPQMP 69

Db	15	LIPFLSLLLPLTPKSAFAOSEPE--LKLESVIVSHGVRAPTKATQLMQDVTDPDWPTWP	73			
QY	70	VKAGYLTFRGAELVTLMGCGFYGYDFRSGLLA--AGCPAEGGVYAQAADIORTLTGTQOAF	128			
Db	74	VKLGLWLTFRGELLIAVLGHYQQRVLVADGLLTKGCPQGOVAIISDVDERTRKTGEAFA	133			
QY	129	DGVARGCGLTWHQADLKKTDPLFHPVETGVCKLDNAQTDKAIERLGGPLDVTVSQRYAK	188			
Db	134	AGLAPDCAITVHTQADTSFDPFLNPLKTVGCQQLDNANVTDAILSRAGGSIADFTGHRQT	193			
QY	189	PFAQMGVDVLPFAASPYCKSLQQCGKTCDFAFHAANEVNVNKEGKVTLSGFLSALSSTLGE	248			
Db	194	AFRELERVLNFPQSNILCLNREKQDBSCSLTQALPSELKVSAD--NVSLTGAVSLASMLTE	251			
QY	249	IFLLQNAQMPVEVAVORLKAENWVLSLSLNAQNLMAKTPYIARHKGTPELLQQIDTAL	308			
Db	252	IFLLQQAGHPFGWRITDSHQWNTLSLNAQYLLQRTPEVARGRATPJDLILMAAL	311			
QY	309	T---LQLDAQGKLPISAQNRVLEFGGHDITANIAGMLGADWOLPBQPDNTPPGGELVF	365			
Db	312	TPHPSQKQAVGVTLPST---VLFTAGHDTNLNLGGALELNWTLPQPDNTPPGGELVF	367			
QY	366	ELWQNPNDHORVYAVKMFYQTDWLQRNAEKLDLKNPAGIISVAVACGE	414			
Db	368	ERWRLLSDNCWITQVSLVFQTLQQMRKTKPLSL--NTPPGEVKTLTAGE	415			
RESULT 11						
Q7UD08						
ID	Q7UD08	PRELIMINARY;	PRT; 432 AA.			
IC	Q7UD08					
DT	01-OCT-2003 (TreMBLrel. 25, Created)					
DT	01-OCT-2003 (TreMBLrel. 25, Last sequence update)					
DT	01-OCT-2003 (TreMBLrel. 25, Last annotation update)					
DE	Phosphoanhydride phosphorylase; pH 2.5 acid phosphatase.					
GN	APPA OR Si048.					
OS	Shigella flexneri.					
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;					
OC	Enterobacteriaceae; Shigella.					
OX	NCBI_TaxID=623;					
RN	[1]_TaxID=623;					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=2457T / ATCC 700930 / Serotype 2a;					
RX	MEDLINE=22590274; PubMed=12704152;					
RA	Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,					
RA	Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,					
RA	Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,					
RA	Schwartz D.C., Blattner F.R.;					
RT	"Complete genome sequence and comparative genomics of Shigella					
RT	flexneri serotype 2a strain 2457T.";					
RL	Infect. Immun. 71:2775-2786(2003).					
DR	EMBL; AEO16981; AAP16495.1; -.					
SQ	SEQUENCE 432 AA; 47136 MW; 200442239BC6EDFC CRC64;					
Query Match						
Best Local Similarity 47.2%; Pred. No. 2.7e-66;						
Matches 193; Conservative 63; Mismatches 138; Indels 15; Gaps 7;						
QY	13	LVLMLSLGIAITAPVA---ABPSGYTLERVLVLSHGVRSPKTKTQLMNDVTPDKPQWP	69			
Db	5	LIPFLSLLLPLTPKSAFAOSEPE--LKLESVIVSHGVRAPTKATQLMQDVTDPDWPTWP	63			
QY	70	VKAGYLTFRGAELVTLMGCGFYGYDFRSGLLA--AGCPAEGGVYAQAADIORTLTGTQOAF	128			
Db	64	VKLGLWLTFRGELLIAVLGHYQQRVLVADGLLTKGCPQGOVAIISDVDERTRKTGEAFT	123			
QY	129	DGVARGCGLTWHQADLKKTDPLFHPVETGVCKLDNAQTDKAIERLGGPLDVTVSQRYAK	188			
Db	124	AGLAPDCAITVHTQADTSFDPFLNPLKTVGCQQLDNANVTDAILSRAGGSIADFTGHRQT	183			
QY	189	PFAQMGVDVLPFAASPYCKSLQQCGKTCDFAFHAANEVNVNKEGKVTLSGFLSALSSTLGE	248			

184 VFRELRLVNFQSNLCLNREKQSCSLTQALPSELKVSAD--NVSLTGAVSLASMLTE 241
249 IFLIQAQMPVAVQWOLKGAENWVSLSLHNAQFNLMKATPVYASHKGTPLQLQDITDAL 308
242 IFLIQAQMPVAVQWOLKGAENWVSLSLHNAQFNLMKATPVYASHKGTPLQLQDITDAL 301
309 T---LQIDAQCKLPISAQNRVLFGLGHDTNINIANIAGLADWOLPEQDNTPPGGGLVP 365
302 TPFPKQKAYGVTLPTS-----VLFIAGHDTNLANLGGALELNWTLPGQDNTPPGGGLVP 357
366 ELWQPNDRHQRVAVKMFYQTMQDLRQNAEKLDLKNPAGIISVAVAGCE 414
358 ERWRRLSDNSQWQVSLVFTQLOQRDKTPLSL--NTPPGVVKLTLAGCE 405

RESULT 12

Q83RW2 PRELIMINARY; PRT; 442 AA.

ID Q83RW2 AC Q83RW2
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Periplasmic glucosylase, pH 2.5 acid phosphatase,
DE periplasmic
GN AFPA OR SF0982.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=301; Serotype 2a; PubMed=12384590;
RX MEDLINE=2227406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao Y., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157.";
RL Nucleic Acids Res. 30:4432-4441 (2002).
DR EMBL; A015127; F:acid phosphatase activity; IEA.
DR GO; GO:0003993; F:acid phosphatase activity; IEA.
DR InterPro; IPR000560; HisAc_phsphtse.
DR Pfam; PF00328; acid phosphat; 1.
DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
DR Complete proteome.
KW Complete proteome.
SQ SEQUENCE 442 AA; 48183 MW; 20BCE2C454254680 CRC64;

Query Match 41.8%; Score 919.5; DB 16; Length 442;
Best Local Similarity 47.2%; Pred. No. 2.8e-66;
Matches 193; Conservative 63; Mismatches 138; Indels 15; Gaps 7;

QY 13 LVLMISGLAATAPVA---AEPGTYLRLVILSHGVRSPTKQTLQMDVTPDKWPWP 69
DB 15 LIPFLISLLIPLTPQSAFAQSEPE-LKLESMTVSRHGVRAPTKATQLMQDTPDAWPTWP 73

QY 70 VKAGYLTTPRGAELVTLMGFGYDYSFSLGLLA-AGCPAEGGYAQAQDIDQRTLTGQAF 128
DB 74 VKLGWLTTPRGGELIIVLGHYQRLVADGLLAKGCPQSAQVAIADVDERTKTGEAFA 133

QY 129 DGVAPOCGLTVHQAQDKKTDPLFHPVETGVCKLNAQDPAKIAERLGGPLDVTQVAK 188
DB 134 AGLAPDCAITVHTQADTSSPDPLNPLKTVGVQQLDNANVTDAILCRAGSIADFTGHRQT 193

QY 189 PFAQMGDLNFAASPYCKSLQOQKTCDFAHFAANEVNVNKGSKVTLISGLPLASLTGCE 248
DB 194 VFRELRLVNFQSNLCLNREKQSCSLTQALPSELKVSAD--NVSLTGAVSLASMLTE 251

QY 249 IFLIQAQMPVAVQWOLKGAENWVSLSLHNAQFNLMKATPVYASHKGTPLQLQDITDAL 308
DB 252 IFLIQAQMPVAVQWOLKGAENWVSLSLHNAQFNLMKATPVYASHKGTPLQLQDITDAL 311

309 T---LQIDAQCKLPISAQNRVLFGLGHDTNINIANIAGLADWOLPEQDNTPPGGGLVP 365
312 TPFPKQKAYGVTLPTS-----VLFIAGHDTNLANLGGALELNWTLPGQDNTPPGGGLVP 367
366 ELWQPNDRHQRVAVKMFYQTMQDLRQNAEKLDLKNPAGIISVAVAGCE 414
368 ERWRRLSDNSQWQVSLVFTQLOQRDKTPLSL--NTPPGVVKLTLAGCE 415

RESULT 13

Q8XBZ6 PRELIMINARY; PRT; 413 AA.

ID Q8XBZ6 AC Q8XBZ6
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Periplasmic glucosylase-1-phosphatase.
GN AGP OR Z1421 OR ECS1158.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamocis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RMD 050952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
RA Kunara S., Shiba T., Hattori M., Shingawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22 (2001).
DR EMBL; AP005294; AAG55550.1; -;
DR EMBL; AP005254; BAB34581.1; -;
DR PIR; B85636; B85636.
DR PIR; P90773; P90773.
DR GO; GO:0003993; F:acid phosphatase activity; IEA.
DR InterPro; IPR000560; HisAc_phsphtse.
DR Pfam; PF00328; acid phosphat; 1.
DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
DR Complete proteome.
KW Complete proteome.
SQ SEQUENCE 413 AA; 45640 MW; A4630644EF66FB1A CRC64;

Query Match 24.2%; Score 532; DB 16; Length 413;
Best Local Similarity 31.3%; Pred. No. 9e-35;
Matches 131; Conservative 83; Mismatches 164; Indels 40; Gaps 11;

QY 10 LSGVLMLSGLAATAPVAAPPSGYTLERVLVILSHGVRSPTKQTLQMDVTPDKWPWP 68
DB 10 VAGVVLASNAQAQTV-----PEGYLQQLVLMVSRHNLRAPLANNNGSVLSQSTENKWPW 64

QY 69 PVKAGYLTTPRGAELVTLMGFGYDYSFSLGLLAAG-CPAEGGYAQAQDIDQRTLTGQAF 127
DB 65 DVPGQGLTKGVLEVTMGHMYREWLAEQGVKSGCEGFPDPTVYVANSQRTVATQAFF 124

QY 128 LDGVAPOCGLTVHQAQDKKTDPLFHPVETGVCKLNAQDPAKIAERLGGPLDVTQVAK 187
DB 125 ITGAPPGCDIPVHQQKMGKMTDPTFNPTDSDSAFSEQVAAMEKELS-----KLQLT 178

Job time : 41.0488 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 30, 2004, 12:35:05 ; Search time 15 Seconds
(without alignments)
2629.235 Million cell updates/sec

Title: US-10-021-723B-13

Perfect score: 2153

Sequence: 1 QSBPELKLESVVIVSRHGVR.....CSLAGFTQIVNEARIPACSL 410

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78.*

1: Piri.*

2: Piri2.*

3: Piri3.*

4: Piri4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2077	96.5	432	2 B36733	acid phosphatase (
2	2042	94.8	434	2 H90770	phosphoanhydride p
3	2042	94.8	444	2 D85633	hypothetical prote
4	954	44.3	441	2 ACO201	acid phosphatase (
5	551	25.6	413	2 F90773	periplasmic glucos
6	551	25.6	413	2 B85636	periplasmic glucos
7	544	25.3	413	2 JV0087	glucose-1-phosphat
8	543	25.2	413	2 AG0632	glucose-1-phosphat
9	534.5	24.8	414	2 E87316	periplasmic phosph
10	518	24.1	417	2 S25627	glucose-1-phosphat
11	133	6.2	416	2 T16058	hypothetical prote
12	132	6.1	423	1 S06167	acid phosphatase (
13	114	5.3	423	2 A33395	acid phosphatase (
14	112.5	5.2	438	2 S84682	acid phosphatase (
15	109.5	5.1	344	2 B89130	protein F52E1.8 [i
16	109.5	5.1	457	2 A56925	paired box transcr
17	109.5	5.1	479	1 JN0890	acid phosphatase (
18	107.5	5.0	459	2 S52250	paired box transcr
19	107.5	5.0	479	1 JN0715	3-phytase (EC 3.1.
20	107	5.0	421	2 S14742	acid phosphatase (
21	107	5.0	5126	2 S40450	ryanodine receptor
22	105	4.9	450	2 A84429	paired box transcr
23	104.5	4.9	397	2 C91716	hypothetical prote
24	103.5	4.8	537	2 S54770	secreted acid phos
25	103.5	4.8	888	2 T46726	secreted acid phos
26	103	4.8	769	2 B97681	tyrosine kinase Di
27	103	4.8	1048	1 BVECSC	exonuclease (EC 3
28	101.5	4.7	356	2 P75594	hypothetical prote
29	101.5	4.7	386	1 JN0610	acid phosphatase (

RESULT 1

B36733

acid phosphatase (EC 3.1.3.2) precursor - Escherichia coli (strain K-12)
N;Alternate names: phosphoanhydride phosphohydrolase, periplasmic; phytase 6
C;Species: Escherichia coli

C;Date: 19-Apr-1991 #sequence revision 19-Apr-1991 #text change 01-Mar-2002

C;Accession: B36733; S18018; B64839; A36534; S17960; S33278

R;Dassa, J.; Marck, C.; Boquet, P.L.

J. Bacteriol. 172, 5497-5500, 1990

A;Title: The complete nucleotide sequence of the Escherichia coli gene appA reveals si

A;Reference number: A36733; MUID:90368616; PMID:2168385

A;Accession: B36733

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-432 <DAS>

A;Cross-references: GB:M58708; NID:9145283; PIDN:AAA72086.1; PID:9145285

R;Greiner, R.; Jany, K.D.

Biol. Chem. Hoppe-Seyler 372, 664-665, 1991

A;Title: Characterization of a phytase from Escherichia coli.

A;Reference number: S18018

A;Accession: S18018

A;Molecule type: protein

A;Residues: 23-33 <GRE>

R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A;Title: The complete genome sequence of Escherichia coli K-12.

A;Reference number: A64720; MUID:97426617; PMID:9278503

A;Accession: B64839

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-432 <BLAT>

A;Cross-references: GB:AE000200; GB:U00096; NID:92367111; PIDN:AA74065.1; PID:9178721

A;Experimental source: strain K-12, substrain MG1655

R;Touati, E.; Danchin, A.

Biochimie 69, 215-221, 1987

A;Title: The structure of the promoter and amino terminal region of the pH 2.5 acid ph

A;Reference number: A28534; MUID:87271766; PMID:3038201

A;Accession: A28534

A;Molecule type: DNA

A;Residues: 1-50, 'NAGCHPRMANLAG', 65, 'T', 67-74, 'DV', 77-111, 'S', <TOU>

A;Cross-references: GB:X05471; NID:940925; PIDN:CAA29031.1; PID:940927

R;Dassa, J.; Fsihi, H.; Marck, C.; Dion, M.; Kieffer-Bontemps, M.; Boquet, P.L.

Mol. Gen. Genet. 229, 341-352, 1991

A;Title: A new oxygen-regulated operon in Escherichia coli comprises the genes for a p

A;Reference number: S17958; MUID:92049231; PMID:1658595

A;Accession: S17960

A;Status: not compared with conceptual translation

A;Molecule type: DNA

A;Residues: 1-17 <DA2>

A;Cross-references: GB:S63811; NID:9238656; PIDN:AAB20286.1; PID:9238659

R;Greiner, R.; Konietzny, U.; Jany, K.D.

Arch. Biochem. Biophys. 303, 107-113, 1993
A;Title: Purification and characterization of two phytases from *Escherichia coli*.
A;Reference number: S33278; MUID:93256556; PMID:8387749
A;Accession: S33278
A;Molecule type: protein
A;Residues: 23-31, 'A', 33-34 <GR2>
C;Comment: In addition to CAMP-mediated control, this enzyme is induced when bacterial
naerobic conditions.
C;Genetics:
A;Gene: appA
A;Map position: 25 min
C;Keywords: monomer; periplasmic space; phosphohistidine; phosphoprotein; phosphoric mon
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-112/Product: acid phosphatase #status predicted <MAT>
F;38/Active site: Arg #status predicted
F;39/Active site: His (phosphohistidine intermediate) #status predicted

Query Match 96.5%; Score 2077; DB 2; Length 432;
Best Local Similarity 99.0%; Pred. No. 4.7e-160;
Matches 402; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 1 QSEPELKESVIVSRHGVRAPTKATQMODVTPDAMPVTPVKGELTPRGGLIAYLGH 60
DB 23 QSEPELKESVIVSRHGVRAPTKATQMODVTPDAMPVTPVKGELTPRGGLIAYLGH 82
QY 61 YWRQRLVADGLLPGKCGPQSGQVVAIIADVDERTKGTGEAFAAGLAPDCAITVHTQADTSS 120
DB 83 YQORQLVADGLLTKKGCPOGQVVAIIADVDERTKGTGEAFAAGLAPDCAITVHTQADTSS 142
QY 121 PDPLFNLKTVGVCOLDNANVTDAILERAGGSIAADFTGHYQTAFARELERVLPQSNLCLK 180
DB 143 PDPLFNLKTVGVCOLDNANVTDAILSRAGSIADFTGHRQTAFARELERVLPQSNLCLK 202
QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 240
DB 203 REKQDESCSLTQALPSELKVSADNVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 262
QY 241 HOWNTLISLHNAQFDILQRTPEVARSRATPLDLIKTALTPHPPOKQAYGVTLPTSVLFI 300
DB 263 HOWNTLISLHNAQFYLLQRTPEVARSRATPLDLIKTALTPHPPOKQAYGVTLPTSVLFI 322
QY 301 AGHDTNLANLGGALLENWTLPGQPDNTPPGGELVFERWRRLSDNSQWIVSLVFTQLQOM 360
DB 323 AGHDTNLANLGGALLENWTLPGQPDNTPPGGELVFERWRRLSDNSQWIVSLVFTQLQOM 382
QY 361 RDKTPLSLNTPPGKVLTLACCEERNAQGMCSLAGFTQIVNEARIPACSL 410
DB 383 RDKTPLSLNTPPGKVLTLACCEERNAQGMCSLAGFTQIVNEARIPACSL 432

RESULT 2
H90770
phosphoanhydride phosphorylase [imported] - *Escherichia coli* (strain O157:H7, substrain
C;Species: *Escherichia coli*
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
R;Accession: H90770
R;Hayashi, T.; Makino, K.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gen
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: H90770
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-434 <HAY>
A;Cross-references: GB:BA000007; PIDN:BA034559.1; PID:q13360596; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RMD 0509952
C;Genetics:
A;Gene: ECS1136

Query Match 94.8%; Score 2042; DB 2; Length 434;
Best Local Similarity 96.3%; Pred. No. 3.2e-157;
Matches 395; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

QY 1 QSEPELKESVIVSRHGVRAPTKATQMODVTPDAMPVTPVKGELTPRGGLIAYLGH 60
DB 25 EPEPELKESVIVSRHGVRAPTKATQMODVTPDAMPVTPVKGELTPRGGLIAYLGH 84
QY 61 YWRQRLVADGLLPGKCGPQSGQVVAIIADVDERTKGTGEAFAAGLAPDCAITVHTQADTSS 120
DB 85 YQORQLVADGLLTKKGCPOGQVVAIIADVDERTKGTGEAFAAGLAPDCAITVHTQADTSS 144
QY 121 PDPLFNLKTVGVCOLDNANVTDAILERAGGSIAADFTGHYQTAFARELERVLPQSNLCLK 180
DB 145 PDPLFNLKTVGVCOLDNANVTDAILSRAGSIADFTGHRQTAFARELERVLPQSNLCLK 204
QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 240
DB 205 REKQDESCSLTQALPSELKVSADNVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 264
QY 241 HOWNTLISLHNAQFDILQRTPEVARSRATPLDLIKTALTPHPPOKQAYGVTLPTSVLFI 300
DB 265 HOWNTLISLHNAQFYLLQRTPEVARSRATPLDLIKTALTPHPPOKQAYGVTLPTSVLFI 324
QY 301 AGHDTNLANLGGALLENWTLPGQPDNTPPGGELVFERWRRLSDNSQWIVSLVFTQLQOM 360
DB 325 AGHDTNLANLGGALLENWTLPGQPDNTPPGGELVFERWRRLSDNSQWIVSLVFTQLQOM 384
QY 361 RDKTPLSLNTPPGKVLTLACCEERNAQGMCSLAGFTQIVNEARIPACSL 410
DB 385 RDKTPLSLNTPPGKVLTLACCEERNAQGMCSLAGFTQIVNEARIPACSL 434

RESULT 3
D85633
hypothetical protein appA [imported] - *Escherichia coli* (strain O157:H7, substrain EDI
C;Species: *Escherichia coli*
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayh
iller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodac
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: D85633
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-444 <STO>
A;Cross-references: GB:AE005174; NID:gi2514245; PIDN:AA055528.1; GSPDB:GN00145; UMGPI:2
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: appA

Query Match 94.8%; Score 2042; DB 2; Length 444;
Best Local Similarity 96.3%; Pred. No. 3.3e-157;
Matches 395; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

QY 1 QSEPELKESVIVSRHGVRAPTKATQMODVTPDAMPVTPVKGELTPRGGLIAYLGH 60
DB 35 EPEPELKESVIVSRHGVRAPTKATQMODVTPDAMPVTPVKGELTPRGGLIAYLGH 94
QY 61 YWRQRLVADGLLPGKCGPQSGQVVAIIADVDERTKGTGEAFAAGLAPDCAITVHTQADTSS 120
DB 95 YQORQLVADGLLTKKGCPOGQVVAIIADVDERTKGTGEAFAAGLAPDCAITVHTQADTSS 154
QY 121 PDPLFNLKTVGVCOLDNANVTDAILERAGGSIAADFTGHYQTAFARELERVLPQSNLCLK 180
DB 155 PDPLFNLKTVGVCOLDNANVTDAILSRAGSIADFTGHRQTAFARELERVLPQSNLCLK 214
QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 240
DB 215 REKQDESCSLTQALPSELKVSADNVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 274
QY 241 HOWNTLISLHNAQFDILQRTPEVARSRATPLDLIKTALTPHPPOKQAYGVTLPTSVLFI 300
DB 275 HOWNTLISLHNAQFYLLQRTPEVARSRATPLDLIKTALTPHPPOKQAYGVTLPTSVLFI 334

QY 301 AGHDTNLANLGALELNTLPGQDNTPTPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 360
 DB 335 AGHDTNLANLGALELNTLPGQDNTPTPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 394
 QY 361 RDKTFLSLNTPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 410
 DB 395 RDKTFLSLNTPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 444

RESULT 4
 AC0201
 acid phosphatase (EC 3.1.3.2) [imported] - Yersinia pestis (strain CO92)
 C:Species: Yersinia pestis
 C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001
 C:Accession: AC0201
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Hoiden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; li, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall, Nature 413, 523-527, 2001
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
 A:Reference number: AB0001; MUID:21470413; PMID:11586360
 A:Accession: AC0201
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-441 <KUR>
 A:Cross-references: GB:AL590842; PIDN:CAC90470.1; PID:gl15979685; GSPDB:GN00175
 A:Gene: YPO1648
 C:Genetics:
 C:Keywords: phosphoric monoester hydrolase

Query Match 44.3%; Score 954; DB 2; Length 441;
 Best Local Similarity 46.8%; Pred. No. 2.9e-69;
 Matches 195; Conservative 65; Mismatches 145; Indels 12; Gaps 6;
 QY 2 SEPE-LKLESVIVSRHGVRAPTKATQMLQMDVTPDAPWTPVKLGELTPRGGLIAYLGH 60
 DB 29 ABSPGYTLERVVLSRHGVRSPTKQTLQMDVTPDAPWTPVKLGELTPRGGLIAYLGH 88
 QY 61 YRQRLVADGLPKKCGPQSGQVIAIADVDERTKGTGEAFAGLAPDCAITVHTQADTSS 120
 DB 89 FYGDFRSGLL-AAGCPAGGVYAGADIDQRLTGQAFDGVAPGGLTVHQAADLKK 147
 QY 121 PDPLFNPLKTGVCOLDNANVTDAILERAGGSIAFTGHYQTAFARELVNPPQSNCLK 180
 DB 148 TDPLFHPVAGVCKLDAQAQTKAIEBQLGGLDTSQRYAKPFAQMGVLYNFAAFSPYCKS 207
 QY 181 REKQDESCSLQALPSELKVSAD-CVSLTGAVSLASMLTEIFLLQQAQMGPEPCWGGRIT 238
 DB 208 LQQQGGTKDFAHFAANEVNVNKGTVLUSGLAUSSTLGEIFLQNAQMEVAVQRLK 267
 QY 239 DSHQWNTLLSLHNAQFDLLQRTPEVARSRATFLDLIKTALTTPHPKQAYGVTLPTS-- 296
 DB 268 GAENWVSLLSLHNAQFNLAQTPYIARHKGTPLLQIDTALT--LQLDAQOKLPISAQ 324
 QY 297 --VLFTAGHDTNLANLGALELNTLPGQDNTPTPGGELVFERWRRLSDNSQWIOVSLVF 354
 DB 325 NRVFLTGGHDTNLANLGALELNTLPGQDNTPTPGGELVFERWRRLSDNSQWIOVSLVF 384
 QY 355 QTLQOMEDKTFSL-NTPPGVEKLTLAGCEERNAQMGCSLAGTQIVNEARIPACSL 410
 DB 385 QTMQDLNRAEKLDIKNNPAGIISVAVAGCENNGDKLCILDITFQKKVAVIRPACHI 441

RESULT 5
 F90773
 periplasmic glucose-1-phosphatase [imported] - Escherichia coli (strain O157:H7, substra
 C:Species: Escherichia coli
 C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
 C:Accession: F90773
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
 Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and ge
 A:Reference number: A99629; MUID:21156231; PMID:11258796
 A:Accession: F90773
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-413 <HAY>
 A:Cross-references: GB:BA000007; PIDN:BA034581.1; PID:gl13360618; GSPDB:GN00154
 A:Experimental source: strain O157:H7, substrain RMD 0509952
 C:Genetics:
 A:Gene: ECs1158

Query Match 25.6%; Score 551; DB 2; Length 413;
 Best Local Similarity 33.3%; Pred. No. 1e-36;
 Matches 138; Conservative 71; Mismatches 169; Indels 36; Gaps 12;
 QY 1 QSEPE-LKLESVIVSRHGVRAPTKATQMLQMDVTPDAPWTPVKLGELTPRGGLIAYL 58
 DB 23 QTVPEGYQLQVLMMSRHLNRLAPLANNQSVLSQSTPNKPEWDVPGGQTLTKGGVLEVYM 82
 QY 59 GHYWRQLVADGLPKKCGPQSGQVIAIADVDERTKGTGEAFAGLAPDCAITVHTQADT 118
 DB 83 GHYWRQLVADGLPKKCGPQSGQVIAIADVDERTKGTGEAFAGLAPDCAITVHTQADT 142
 QY 119 SSPDPLFNPLKTGVCOLDNANVTDAILERAGGSIAFTGHYQTAFARELVNPPQSNCLK 175
 DB 143 GTWDPFTFNFTD---DAAAFSEQAAWAMEKLSKL-QLTDSYQL-----LEKIVNKDS 193
 QY 176 NLCKEKKDESCSLQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMGPE--EPG 233
 DB 194 PAC-----XKQQCSLVGDKNTFSKAKYQBPVSGVGLKVGNSLVDAFTLQVYEGFPMDQVA 249
 QY 234 WGRITDSHONTLLSLHNAQFDLLQRTPEVARSRATFLDLIKTALTTPHPKQAYGVTL 293
 DB 250 WGBIKSDQWQKLSKLGYSQSLFTSPEVARNVAKPLVSYIDKAL-----VTD 298
 QY 294 PTS---VLFTAGHDTNLANLGALELNTLPGQDNTPTPGGELVFERWRRLSDNSQWIO 349
 DB 299 RTSAPKITVLVGHDSNIALSLTALDPKPYQLQHDNERTPIGGKIVFORWEDSKANRDLMK 358
 QY 350 VSLVFQTLQOMEDKTFSL-NTPPGVEKLTLAGCEERNAQMGCSLAGTQIVNEA 403
 DB 359 IEVYQSAEQLRNADALTLOAPQRTLELSGC-PIDADGFCPMDKFPDSVLNFA 411

RESULT 6
 B95636
 periplasmic glucose-1-phosphatase [imported] - Escherichia coli (strain O157:H7, subse
 C:Species: Escherichia coli
 C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: B95636
 R:Perera, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayh
 iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamou, K.; Apodac
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: B95636
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-413 <STO>
 A:Cross-references: GB:AE005174; NID:gl21514273; PIDN:AAG55550.1; GSPDB:GN00145; UWGP:Z
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: agp

Query Match 25.6%; Score 551; DB 2; Length 413;
 Best Local Similarity 33.3%; Pred. No. 1e-36;
 Matches 138; Conservative 71; Mismatches 169; Indels 36; Gaps 12;
 QY 1 QSEPE-LKLESVIVSRHGVRAPTKATQMLQMDVTPDAPWTPVKLGELTPRGGLIAYL 58
 DB 23 QTVPEGYQLQVLMMSRHLNRLAPLANNQSVLSQSTPNKPEWDVPGGQTLTKGGVLEVYM 82
 QY 59 GHYWRQLVADGLPKKCGPQSGQVIAIADVDERTKGTGEAFAGLAPDCAITVHTQADT 118

Db 83 GHYREWLAEQGVKSGCPEDTAVYANSIQRVATAQFFITGAFGCCDIPVHQBKM 142
QY 119 SSPDLFNLKGTGVCOLDNANVTD---AILERAGGSIAFTGHYQTAFARELERVLPPOS 175
Db 143 GTMDPTFNPVITD---DSAAFSEQVAAMEKLSKL-QLTDSYQL-----LEKIVNYKDS 193
QY 176 NLCKREKQDESGSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLQQAQGM--EPG 233
Db 194 PAC-----KEKQCSLVGKNTSFAKYQQQEGVSGPLKVGNSLVDAFTLYIEGFPMDQVA 249
QY 234 WGRITDHWNTLLSHNAQFDLLQRTPEVARSRAFPDLDLIKTALTTPHPKQAYGVTL 293
Db 250 WGEIKSDQWKVLSKLGKNGYQDSLFTSPEVARNVAKPLVSYIDKAL-----VTD 298
QY 294 PTS---VLFITAGHTNLNLANGGALELN-WTLPGQPDNTPPGGELVFERWRRLSDNSQWIO 349
Db 299 RTSAPKITVLVGHDSNIASLLTALDFKPYQLHDQNERFTIGGKIVFORWHDGSKANRDLMK 358
QY 350 VSLVFQTLQOMRDKTFLSLNTPPGVEVKLTLAGCEERNAQMGCSLAGFTQIVNEA 403
Db 359 IEVYQSAEQLRNADALTLOAPQRVTLLESGC-PIDADGFCPMDKFDVSLNEA 411

RESULT 7

JV0087
Glucose-1-phosphatase (EC 3.1.3.10) precursor - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 01-Mar-2002
C:Accession: JV0087; H64841
C:Fraser, E.; Marck, C.; Boquet, P.L.
J. Bacteriol. 172, 802-807, 1990
A:Title: Nucleotide sequence and transcriptional analysis of the Escherichia coli agp gene
A:Reference number: JV0087; MUID:90130318; PMID:2153660
A:Accession: JV0087
A:Molecule type: DNA
A:Residues: 1-413 <PRA>
A:Cross-references: GS:M33807; NID:G145217; PIDN:AAA23426.1; PID:G145218
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: H64841
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-413 <BLAT>
A:Cross-references: GB:AE000202; GB:U00096; NID:G1787233; PIDN:AAC74087.1; PID:G1787237;
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: agp
A:Map position: 23 min
C:Function:
A:Description: essential for growth in a high-phosphate medium containing glucose-1-phosphate
A:Note: optimal at low pH
C:Keywords: homodimer; periplasmic space; phosphohistidine; phosphoprotein; phosphoric acid
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-413/Product: Glucose-1-phosphatase #status predicted <MAT>
F:40/Active site: His (phosphohistidine intermediate) #status predicted
F:311/Active site: His #status predicted

Query Match 25.3%; Score 544; DB 2; Length 413;
Best Local Similarity 33.1%; Pred. No. 3.7e-36;
Matches 137; Conservative 71; Mismatches 170; Indels 36; Gaps 12;

QY 1 QSEPE-LKLESVIVSRHGVRAP-TKATQLMODVTPDAMPWPVKLGELTPRGGLIAYL 58
Db 23 QTPPEGYQLQVLMMSRNLRAPLANNVLAQSTPNKWPNDVPGGQLTTKGGVLEYVM 82
QY 59 GHYWRQLVADGLPKCGCPQSGQVAIIADVDERTRKTGEAFAAGLAPDCAITVHTQADT 118
Db 83 GHYREWLAEQGVKSGCPEPTAVYANSIQRVATAQFFITGAFGCCDIPVHQBKM 142

QY 119 SSPDLFNLKGTGVCOLDNANVTD---AILERAGGSIAFTGHYQTAFARELERVLPPOS 175
Db 143 GTMDPTFNPVITD---DSAAFSEQVAAMEKLSKL-QLTDSYQL-----LEKIVNYKDS 193
QY 176 NLCKREKQDESGSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLQQAQGM--EPG 233
Db 194 PAC-----KEKQCSLVGKNTSFAKYQQQEGVSGPLKVGNSLVDAFTLYIEGFPMDQVA 249
QY 234 WGRITDHWNTLLSHNAQFDLLQRTPEVARSRAFPDLDLIKTALTTPHPKQAYGVTL 293
Db 250 WGEIKSDQWKVLSKLGKNGYQDSLFTSPEVARNVAKPLVSYIDKAL-----VTD 298
QY 294 PTS---VLFITAGHTNLNLANGGALELN-WTLPGQPDNTPPGGELVFERWRRLSDNSQWIO 349
Db 299 RTSAPKITVLVGHDSNIASLLTALDFKPYQLHDQNERFTIGGKIVFORWHDGSKANRDLMK 358
QY 350 VSLVFQTLQOMRDKTFLSLNTPPGVEVKLTLAGCEERNAQMGCSLAGFTQIVNEA 403
Db 359 IEVYQSAEQLRNADALTLOAPQRVTLLESGC-PIDADGFCPMDKFDVSLNEA 411

RESULT 8

AG0632
Glucose-1-phosphatase precursor (GIPase), secreted [imported] - Salmonella enterica serovar Typhimurium
C:Species: Salmonella enterica subsp. enterica serovar Typhimurium
A:Note: this species has also been called Salmonella typhimurium
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AG0632
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Park, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AG0632
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-413 <PAR>
A:Cross-references: GS:AL513382; PIDN:CAD08242.1; PID:G16502289; GSPDB:GN00176
C:Genetics:
A:Gene: STY1153

Query Match 25.2%; Score 543; DB 2; Length 413;
Best Local Similarity 33.6%; Pred. No. 4.5e-36;
Matches 137; Conservative 66; Mismatches 181; Indels 24; Gaps 10;

QY 1 QSEPE-LKLESVIVSRHGVRAP-TKATQLMODVTPDAMPWPVKLGELTPRGGLIAYL 58
Db 23 QTPPEGYQLQVLMMSRNLRAPLANNVLAQSTPNKWPNDVPGGQLTTKGGVLEYVM 82
QY 59 GHYWRQLVADGLPKCGCPQSGQVAIIADVDERTRKTGEAFAAGLAPDCAITVHTQADT 118
Db 83 GHYREWLAEQGLIPSGCPADPTVYVANSIQRVATAQFFITSAFPGCDIPVHQBKM 142
QY 119 SSPDLFNLKGTGVCOLDNANVTDAILERAGGSIAFTGHYQTAFARELERVLPPOSNLC 178
Db 143 GTMDPTFNPVITDSSAAFQQAQVA-MEKARSQ-----HLDSEYKLEIQITHYQDSPEC 196
QY 179 LKREKQDESGSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLQQAQGM--EPGWR 236
Db 197 -----KEKQCSLIDAKTFSANYQQQEPGQGLKVGNSLVDAFTLYIEGFPMDQVWGG 252
QY 237 ITDHWNTLLSHNAQFDLLQRTPEVARSRAFPDLDLIKTALTTPHPKQAYGVTLPTS 296
Db 253 IHTDQWKVLSKLGKNGYQDSLFTSPTVARNVAAPIVKYIDKVLV-----AERVSAP-K 304
QY 297 VLFITAGHTNLNLANGGALELN-WTLPGQPDNTPPGGELVFERWRRLSDNSQWIO 355
Db 305 VTLVGHDSNIASLLTALDFKPYQLHDQYERTPIGGQLVFORWHDGSKANRDLMKTEYVQ 364
QY 356 TLOOMRDKTFLSLNTPPGVEVKLTLAGCEERNAQMGCSLAGFTQIVNEA 403

Db 365 SARQLRNAEALTLKSPAQRVTLKGCPCV-DANGFCPLDKFDNMNTA 411
RESULT 9
E87316
periplasmic phosphoanhydride phosphohydrolase [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: E87316
R;Niernman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of *Caulobacter crescentus*.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: E87316
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-414 <STO>
A:Cross-references: GB:AE005673; NID:g13421731; PIDN:AAK22529.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC0542

Query Match 24.8%; Score 534.5; DB 2; Length 414;
Best Local Similarity 33.7%; Pred. No. 2.2e-35;
Matches 137; Conservative 60; Mismatches 177; Indels 33; Gaps 9;
QY 8 LESVIVSRHGVRAPTKATQMLQMDVTPDAMPTVPVKLGELTPRGGLIAYLGHYRQLV 67
Db 25 LEKVILSRHGVRSMSPERLEASARPWERFEVPAGHLTARGETLVASMGDYRHYA 84
QY 68 ADGLLPKCGCQSQVAIIADVDRTRKTGEAPAGLAPDCAITVHTQADTSPDPLFNP 127
Db 85 AQGLLPKCGDC---ASVYAWANVTQRTATAKAYRETLAPGCPVTVTVGE-GNIDPMFEP 140
QY 128 LKTVGCOLDNANVTDAILERAGGSIADFTGHYQTAFARELERVLPQSNCLKREKQDES 187
Db 141 VKAGIVKADHALARAANVAGRVGGDLTAWASHNQEAQLDALLMQCDKGPC----- 191
QY 188 CSLTQALPSELUK-----SADCVSLTGAVSLASMLTEIFLLQQAQMPGCG-RI 237
Db 192 ----PPAPGKRVRFDKAFPGFYDGBELAGLSGFEAFAGVTESLMAWADGRDFAGLGWKS 247
QY 238 TDSHQWNTLLSHNAQFDLQRTPEVARSRATPLDLIKTALTTPHPKQAYGVTLPSTV 297
Db 248 LDEALTRSFLLHQAEPLRLRTTYPVATLGLADRLAATLR---DGAAAG-PVDARL 303
QY 298 LFIAGHDTNLANLGGALELNLTPG-QPDNTPPGGELVFERWRRLSDNSQWISLVFQT 356
Db 304 VTIAGHDGTLASLGGLLRLMWTLPYQPNQIQPGGALVFERWR-DDGVRVVRVFTGQS 362
QY 357 LQOMRDKTPLSLNTPPGEVKTLTACCEBNAQGCAGTQIVNEA 403
Db 363 LSQRNMTALDAKTPPLSAPVFOGCGTATPAFCRLDEDFETVVRGA 409

RESULT 10
S25627
Glucose-1-phosphatase (EC 3.1.3.10) precursor - Providencia rettgeri
C:Species: Providencia rettgeri
C:Date: 25-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 08-Oct-1999
C:Accession: S25627
R;Riccio, M.L.; Chiesurin, A.; Lombardi, G.; Satta, G.
submitted to the EMBL Data Library, September 1992
A:Reference number: S25627
A:Accession: S25627
A:Molecule type: DNA
A:Residues: 1-417 <RIC>
A:Cross-references: EMBL:X68201; NID:g45771; PIDN:CAA46288.1; PID:g45772
A:Experimental source: strain PV7
C:Genetics:
A:Gene: ggp
C:Keywords: periplasmic space; phosphohistidine; phosphoprotein; phosphoric monoester by

F1-29/Domain: signal sequence #status predicted <SIG>
F130-417/Product: glucose-1-phosphatase #status predicted <MAT>
F142/Active site: His (phosphohistidine intermediate) #status predicted
F1314/Active site: His #status predicted

Query Match 24.1%; Score 518; DB 2; Length 417;
Best Local Similarity 30.4%; Pred. No. 4.7e-34;
Matches 126; Conservative 83; Mismatches 167; Indels 38; Gaps 11;
QY 2 SEPELKLSSVIVSRHGVRAPTKATQMLQMDVTPDAMPTVPVKLGELTPRGGLIAYLGHY 61
Db 27 NQADWLVLDQVLVLSHNRURFIVNTGILTEVTDKKWPDMDAKSGVLTITQGGALEVYMGHY 86
QY 62 WRQ-----RLVADGLLPKCGCQSQQ-VAIADVDRTRKTGEAPAGLAPDCAITVHTQ 115
Db 87 FREWIDQNKLADEL-----CPTSNEDIYLYTNSLQRTIATAQFPAAAGPCGKVNIIHQ 141
QY 116 ADTSSPDPLFNPPLKTVGVCOLDNANVTDAILERAGGSIADFTG---HYQTAFRELERVLP 173
Db 142 PEIGKNDPVFNPIIT-----NGSPFKQKALAMDDYLGSLKAGYBELDTVLNIK 193
QY 174 QSNCLKREKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMP--E 231
Db 194 DSOKC---KTDKLCNLDQSKNSFIIEADKEPVGSGPLKIANSAVDAIDLQYIEGFPADQ 249
QY 232 POGWBITDSHOWNTLLSHNAQFDLQRTPEVARSRATPLDLIKTALTTPHPKQAYGV 291
Db 250 VAWGLVDTPFERKWKLNLTAKNAQETLTPKTIKNAVAHPILNYDKGVFVSDKGETA--- 306
QY 292 TLPTSVLFTAGHDTNLANLGGALELN-WLPQPDNTPPGGELVFERWRRLSDNSQWITQV 350
Db 307 ----KFIFLVGHDSNIASIMSAMDPKYQLAQAYEHTPIGGKLVFORWTDQTKKDFMKV 362
QY 351 SLVFTQLQOMRDKTPLSLNTPPGEVKTLTACGE-ERNAQGCAGTQIVNEA 403
Db 363 EYVYQTADQLRDNAYLSLETPEKHTVLELKDCPVDKN--GYCSWEDFOKWMKA 414

RESULT 11
T16058
hypothetical protein F13D11.1 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 23-Mar-2001
C:Accession: T16058
R;Fulton, L.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of *C. elegans* cosmid F13D11.
A:Reference number: S69020
A:Accession: T16058
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-416 <FUL>
A:Cross-references: EMBL:U40939; NID:g1072175; PID:g1072178; PIDN:AAA81702.1; CESP:F131
C:Genetics:
A:Gene: CESP:F13D11.1
A:Introns: 23/3; 88/3; 138/3; 202/3; 244/1; 270/3; 312/3; 349/3; 377/1
C:Superfamily: mammalian acid phosphatase

Query Match 6.2%; Score 133; DB 2; Length 416;
Best Local Similarity 21.5%; Pred. No. 0.0063;
Matches 92; Conservative 57; Mismatches 161; Indels 118; Gaps 18;
QY 7 KLESVIVSRHGVRAPTKATQMLQMDVTPD-----AWPTVPVKLGELTPRGGLIAYLGH 60
Db 17 KLFVQVIFRFGERTFGHYLPFGDDLNNVDYQQTAWP-----GELTKRGILERFQLG- 69
QY 61 YMRQL-----VADGLLPKCGCQSQGVAIADVDRTRKTGEAPAGLAPDCAITVH 113
Db 70 ---QRLKRYGEFGTGYPR-----DFHYVTGKDNITSASAQAMFAGFLPP----- 113
QY 114 TQADTSPDPLFNPPLKTVGVCOLDNANVTDAILERAGGSIADFTGHYQTAFARE-----L 166
Db 114 NEDQTNWYELKQFPV-----AQLTDESIDWVSLGAIIDNCPVYGEAQKSESSEYAEVM 164

```

167  ERLVNPPOSNLCRKREKODESCSLTQALPSELKVSADCVSLTGAVSLASMLTETI---FLL 223
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
165  DOMEXYDAELLQLVFNHADE-----PIVSAVKYNHVIDSLVKRYTIL 205
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
224  QOACQWPEPGWGRIRITDSHQWNTLLSLHNAQFDLIQRTPEVARSRA TELLD---LIK TAL 279
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
206  QDDR-LPYPEWARGYENRILNWSFLI HDA-----VVKQNDSVG DYHNELVMSYF 254
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
280  TTHPPQKQAYGVTLPTSVLFTAGHDTNLANLGGALELNWTLPGQONTTP-GGELVFERW 338
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
255  ETHLOKNSTKV-----FISGHDTNLVTIWESRLD---GHPEIDPNYGAHTAIE-- 301
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
339  RELSDNSQWIOVSLVFQTLQOMRDKTPLSLNTPGPEVKLTLAGCEBNAQCMCSLAGFTQ 398
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
302  -----MHEPVGQLSIKFLLSMGQNTRVELHPHC-SRSQNNDC TWDDEFOR 346
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
399  IVNEARIP 406
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
347  LVKSKRKP 354
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 12

acid phosphatase (EC 3.1.3.2) ACP2 precursor [validated] - human
N/Alternate names: acid phosphatase, lysosomal
C/Species: Homo sapiens (man)
C/Date: 04-Dec-1992 #sequence revision 01-Dec-1995 #text_change 08-Dec-2000
C/Accession: S06167; S05525; S01155
R/von Figura, K.
submitted to the EMBL Data Library, June 1989
A/Reference number: S06167
A/Accession: S06167
A/Molecule type: DNA
A/Residues: 1-423 <VON>
A/Cross-references: EMBL:X15525; NID:G34239; PID:CAA33542.1; PID:G1199524
R/Geier, C.; von Figura, K.; Pohlmann, R.
Eur. J. Biochem. 183: 611-616, 1989
A/Title: Structure of the human lysosomal acid phosphatase gene.
A/Reference number: S05525; MUID:8937828; PMID:2776754
A/Accession: S05525
A/Molecule type: DNA
A/Residues: 1-29 <GEI>
R/Rohlmann, R.; Krentler, C.; Schmidt, B.; Schroeder, W.; Lorkowski, G.; Cui, C.
EMBO J. 7: 2343-2350, 1988
A/Title: Human lysosomal acid phosphatase: cloning, expression and chromosomal localization.
A/Reference number: S01155; MUID:89052645; PMID:3191910
A/Accession: S01155
A/Molecule type: mRNA
A/Residues: 1-423 <POH>
A/Cross-references: EMBL:X12548; NID:G34262; PID:CAA31064.1; PID:G34263
A/Note: part of this sequence, including the amino end of the mature protein.
C/Genetics:
A/Gene: GDB:ACP2
A/Cross-references: GDB:118963; OMIM:171650
A/Map position: lip1.2-lip1.11
A/Introns: 38/3; 70/3; 99/3; 150/3; 183/3; 213/3; 258/1; 285/3; 321/2; 380/3
C/Function:
A/Description: catalyzes the hydrolysis of a wide range of phosphate esters
C/Superfamily: mammalian acid phosphatase
C/Keywords: glycoprotein; phosphatidine; phosphoprotein; phosphoric monoester
F/1-30/Domain: signal sequence #status predicted <SIG>
F/31-423/Product: acid phosphatase ACP2 #status experimental <MAT>
F/42/Active site: Arg #status predicted
F/42/Active site: His (phosphatidate intermediate) #status predicted
F/92,133,167,177,191,267,322,331/Binding site: carbohydrate (Asn) (covalent bond) #status predicted
F/159-370,212-310,345-349/Disulfide bonds: #status predicted
F/159-370,212-310,345-349/Disulfide bonds: #status predicted

Query Match	6.1%	Score 132;	DB 1;	Length 423;
Best Local Similarity	23.3%	Pred. No. 0.0077;		

Best Local Similarity 25.3%, Fied. No. 0.0077
Matches 97; Conservative 43; Mismatches 173; Indels 104; Gaps 20;
QY 8 LESWIVSRHGVRAPTKATQLMODVTDPAAWTPVKLGELTPRGELIAYLGHYWRRLV 67

	: : : : : : :	33 LRFTVLLYRHDRGSFVKT-----YPKDPYOBEWPFQGLTKEGMLQHWELOAURQRY- 87
D b		
	: : : : : : :	68 ADGILLPKCGPQSOGVAIIADV-D-ERRRTKGFAAGLAPDCAITVHTOADTSPPDL-- 124
Q y		
	: : : : : : :	88 -HGFL---NTSYHRQEYVVVRSTDPRTLMSAEANLAGLP-----PNGMQR 129
D b		
	: : : : : : :	125 FNP-LKTGVCOLDNNAVTDAILERAGGSIAFDTHGYQTAFRELVLNBPQSNCLIKRK 183
Q y		
	: : : : : : :	130 ENFNISWQPIFVHVPITE-----DRLKLPLGPCPYEQL 165
D b		
	: : : : : : :	184 QDESCLSICALPSBLKYSDCVSLTGAVLSASMLTEIFL-----LQAQMGPSP 232
Q y		
	: : : : : : :	166 QNE----TRQPPEVCNESSRNAOFLDMVANETGLDTLETWNVDYTLFCETHGLRIP 221
D b		
	: : : : : : :	233 GNGRITDSHOWNTLLSHNAQDLL---ORTPVEVASRATPLLCLKLTALTTHPPKKQAY 289
Q y		
	: : : : : : :	222 PW---ASPQTMRLSRIADFSRELFGIYQOAELQGGLLAQIRKNLTLMATTSQ-- 276
D b		
	: : : : : : :	290 GWTLPTSVLFIAGHDNTLANLGAALENWTLPQQPDNTPPGGBLVFERWRRLSDNSOWIQ 349
Q y		
	: : : : : : :	277 ----LP-KULVTYSAHDTTVIALQMALDY-----NGEQAPYASCHIELFYQEDSEN---RS 324
D b		
	: : : : : : :	350 VSLVFQTTQQMRDKTPTSLNTPPGSVKTLTAGCEERNAQCMSGLAGTOIVNEARIP 406
Q y		
	: : : : : : :	325 VMYYE---RNESDKAPPLSLP-----GCPHR-----CPLQDFLRLETPPVVP 363
D b		

RESULT 13

A33395
acid phosphatase (EC 3.1.3.2) precursor - rat
C/Species: Rattus norvegicus (Norway rat)
C/Date: 29-Jan-1990 #sequence_revision 29-Jan-1990 #text_change 22-Jun-1999
C/Accession: A33395
R:Himeno, M.; Fujita, H.; Noguchi, Y.; Kono, A.; Kato, K.
Biochem. Biophys. Res. Commun. 162, 1044-1053, 1989
A>Title: Isolation and sequencing of a cDNA clone encoding acid phosphatase in rat liv
A/Reference number: A33395; MUID:89350910; PMID:2764916
A/Accession: A33395
A>Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-423 #HM>
A/Cross-references: GB:M27893, MID:G202933; PIDN:AAA0744.1; PID:G202934
C/Superfamily: mammalian acid phosphatase
C/Keywords: phosphohistidine; phosphoprotein; phosphoric monoester hydrolase
F/41/Active site: Arg #status predicted
F/42/Active site: His (phosphohistidine intermediate) #status predicted

Query Match 5.3%: Score 114; DB 2; Length 423;

Query Match	5.3%	Score 114	DB 2	Length 423
Best Local Similarity	22.7%	Pred. No. 0.22		
Matches	95	Conservative	42	Mismatches 175
Indels	106	Gaps	20	

Qy	8	LSWVIVSRHGVRAPTKATQLMQOVTDPAWPTWPKUGELTPRGGELIAYLGHYWRQRLV	67
Db	33	LRFTVLLYRHGDRSPVKA-----YPKDPYQEBEKKPQGFQGLTKGMLGHWSLQALHQRV-	87
Qy	68	ADGLLPKCGCPOSGQGVAAIADV-ERRTKTGEAPAAAGLAPDCAITVHTQADTSSPPLEN	126
Db	88	HGFL---NASVHRGEVYVRSTDFTRLMSAEANLAGLFPF-----TEVCHFNPNISMQ	137
Qy	127	PKTGVQCOLDNANVTDAILERAGGSIADFTGHYQTARELERVLNTPQSNCLKREKODE	166
Db	138	PI-----PVHTVPITE-----DRLLKFPGLGCPRYEQLONE	168
Qy	187	SCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFL-----LQAQGMPSFGWG	235
Db	169	---IRQTEPYQMSIQAQFLDWMANETGLNLTETIWNVYDTLFCETHGLLPPW-223	223
Qy	236	RITDSHOWNTLISLHNAQF-----DLLQETPEVASRATPLDLIKTALTTPHPQKOA	288
Db	224	--ASPQVQALSQLKDFSLFLFGIHQVQK-----ARLQGGVLLAQILKLNLTLMATTSQF	277
Qy	289	YGVTLPTSVLFIAGHTDNLNLAGALELNWTLPCQDPNTPPGGELVFERWRRLSDNSQWI	348

278 -----PKLLVSAHDTTLVALQALNVY-----NGKQAPVASCHIFELVQ--EDNGNF- 323
QY 349 QVSLVFQTLQMRDKTSLNTPGVEKLTLAGCEERNAQGMCSLAGFTQIVNEARIP 406
Db 324 SVEMVFRNDSK---KAPW-----PLTLPGCPHR-----CPLQDFLKL-TEFVIP 363

RESULT 14
S64682
acid phosphatase (EC 3.1.3.2) Acph-1 precursor, lysosomal - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 06-Dec-1996 #sequence_revision 06-Dec-1996 #text_change 17-Mar-2000
C:Accession: S64681
R:Chung, H.J.; Shaffer, C.; MacIntyre, R.
Mol. Gen. Genet. 250, 635-646, 1996
A:Title: Molecular characterization of the lysosomal acid phosphatase from Drosophila melanogaster
A:Reference number: S64681; MUID:96194627; PMID:8676866
A:Accession: S64682
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-438 <CHU>
A:Accession: S64681
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-31, 'Y', 33-438 <CHW>
C:Genetics:
A:Gene: FlyBase:Acph-1
A:Cross-references: FlyBase:FBgn0000032
C:Superfamily: mammalian acid phosphatase
C:Keywords: phosphoric monoester hydrolase
F:1-33/Domain: signal sequence #status predicted <SIG>
F:34-438/Product: acid phosphatase Acph-1 #status predicted <MAT>

Query Match 5.2%; Score 112.5; DB 2; Length 438;
Best Local Similarity 20.3%; Pred. No. 0.31;
Matches 82; Conservative 59; Mismatches 145; Indels 117; Gaps 19;

QY 7 KLESVIVSRHGVRAPTKATQLMQDVTDPDAMPT-----WPVKLGELTPRGGLTAYL 58
Db 47 QLKFWHVYRGRTPV-----DPPTDFWGRKFWPTGWDGLNLGKQEHYDL 95
QY 59 GHYWRQRLVADGLPKGCPGQGVIAIDVD--ERTRKTGEAFAAGL-----APDC----- 108
Db 96 GKMLRNRY--SNLLPPI---YSNENIYVQSTDVDRTLMSAQSLAGLYEPQGEDIWNTDI 150
QY 109 ---AIVHTQADTSSDPDLNPLKTVGCOLDNANVTDAILERAGGSIADFTGHYQTAFRE 165
Db 151 NWQIPIHISPERDDP-----ILAAKAPCPAYDY-----E 180
QY 166 LERVLPQSNLCLKREKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQ 225
Db 181 LASLESSPEFK-ALTEKHNLFAYLSEKGRPVKTFID-----AQYLNLTLEFEN 229
QY 226 AQGWPEPCWGRITDISHQWNTLLSLHNAQFDLLQRTPEVARSRATPLL-DLIKALTALTPHP 284
Db 230 LYNNTLPKWTKVYGREELATYVS--NFAFAISSYTRKLARKAGFLKIDIFORF----- 281
QY 285 QKQAYGVTLPTSLVLF-AGHDNLNLANLGALLENWTLPGQPDNTPPGGELVFERWRLSD 343
Db 282 KEKSSGSLKPDBSMWVYSAHDTTVA SVLNALKL-FEL-----HSPP----- 321
QY 344 NSQWIQVSLVPTQLQMRDKTPL-----SLNTPPGVEKLTLAGC 382
Db 322 -----YTACIMMELRVDETNTPSLVIFYKNTTASPLPLDIPGC 359

RESULT 15
B89130
Protein F52E1.8 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 03-Mar-2003
C:Accession: B89130

R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_e-
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: B89130
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-344 <STO>
A:Cross-references: GB:chr_V; PIDN:AAB37044.1; PID:gl086806; GSPDB:GN00023; CESP:F52E1
C:Genetics:
A:Gene: F52E1.8
A:Map position: 5
C:Superfamily: mammalian acid phosphatase

Query Match 5.1%; Score 109.5; DB 2; Length 344;
Best Local Similarity 19.2%; Pred. No. 0.38;
Matches 82; Conservative 63; Mismatches 138; Indels 145; Gaps 19;

QY 16 RHGVRAPTKATQLMQDVTDPDAMPT--WPVKLGELTPRGGLIAYLGHYWRQRLVADGLLPK 74
Db 6 RHGDRAPGDLPPKDKYNETFWPGW----DQLTKNGIWAQVAVELGILWRQRYGAT-VLP- 59
QY 75 CGCPGSGQVIAIDVDERTKTRKTGEAFAAGLAP--DCAITVHTQADTSSDPDLNPLKTVG 132
Db 60 --IFDKDKVFIILSDSERAETAGQVSAGLFPFVDDRVWESSYLRYWQTPTI----- 109
QY 133 COLDNANVTDAILERAGGSIADFTGHYQTAFRELSRVL-----NFPQSNLCLKREKQ 184
Db 110 -----QTAYGTIDALLRPTKVKCFNYDLAN-----EQ 136
QY 185 DESCSLTQALPSELKVSADCVSLTGAVSLASML-LTEIFLLQQA--QGMPEPGW----- 234
Db 137 EAPAIATQINNEYGQVFNWLNQITGMESIDFWNINDLYDIQRELDHNMFPQWLNQVFN 196
QY 235 -----GRITDISHQWNTLLSLHNAQF-----DLLQRTPEVARSRATPLDLIKT 277
Db 197 TTIMDHIRELKRITRNQEFN---SPTKAFRGGLVNGFLQNMEDLKANKTT----- 245
QY 278 ALTPHPPOKQAYGVTLPTSLVLFIAGHDNLNLANLGALLENWTLPGQPDNTPPGGELVFER 337
Db 246 -----KNAMYSHSDGTL SALLYALNVS-----NDQLVPYVATVLFEL 283
QY 338 WRRLSDNSQWIOVSLVFQTLQMRDKTSLNTPGVEKLTLAGCEERNAQGMCSLAGFT 397
Db 284 Y---DNT---VQLFYK-----NTTSTAYPMTIPGCGQ-----ICPYSQPL 318
QY 398 QIVNEARI 405
Db 319 QLENVRV 326

Search completed: April 30, 2004, 12:42:20
Job time : 17 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 30, 2004, 12:31:10 ; Search time 10 Seconds

(without alignments)
2134.876 Million cell updates/sec

Title: US-10-021-723B-13

Perfect score: 2153

Sequence: 1 QSBPELKLEWVIVSRHGV.....CSLAGFTQIVNEARIPACSL 410

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2077	96.5	432	1	PPA_ECOLI	P07102 escherichia
2	548	25.5	413	1	AGP_SALTY	O33921 salmonella
3	544	25.3	413	1	AGP_ECOLI	P19926 escherichia
4	518	24.1	417	1	AGP_PROSE	O52309 providencia
5	132	6.1	423	1	PPAL_HUMAN	P11117 homo sapien
6	114	5.3	423	1	PPAL_RAT	P20611 rattus norv
7	111.5	5.2	457	1	PXAS_MOUSE	Q00288 mus musculu
8	109.5	5.1	479	1	PHYB_ASPAW	P34755 aspergillus
9	109	5.1	5127	1	RY44_DROME	Q24498 drosophila
10	108.5	5.0	479	1	PHYB_ASPNG	P34754 aspergillus
11	107.5	5.0	459	1	PXAS_CANPA	P47240 canis famil
12	105	4.9	450	1	PXAS_HUMAN	Q06710 homo sapien
13	104	4.8	423	1	PPAL_MOUSE	P24638 mus musculu
14	103	4.8	769	1	DIVL_CAUCR	Q97999 caulobacter
15	103	4.8	1048	1	SBCC_ECOLI	P13458 escherichia
16	101.5	4.7	386	1	PPAP_HUMAN	P15309 homo sapien
17	100	4.6	635	1	TPOR_HUMAN	P40238 homo sapien
18	95.5	4.4	419	1	ARGT_SNVY3	P74122 s arginine
19	94	4.4	341	1	TIN2_MOUSE	Q98299 mus musculu
20	94	4.4	612	1	MCR_XENLA	O91573 xenopus lae
21	94	4.4	895	1	DAGI_BOVIN	O18738 bos taurus
22	93.5	4.3	381	1	PPAP_RAT	P20646 rattus norv
23	92.5	4.3	625	1	TPOR_MOUSE	Q08351 mus musculu
24	92.5	4.3	702	1	CANB_HUMAN	Q9umq6 homo sapien
25	92.5	4.3	1085	1	CASR_BOVIN	P35384 bos taurus
26	91	4.2	437	1	LH2A_HUMAN	O00534 homo sapien
27	91	4.2	458	1	PXAS_RAT	P51974 rattus norv
28	91	4.2	551	1	CAP_SCHPO	P36621 schizosacch
29	91	4.2	3033	1	POLG_HCVJ6	P26650 h genome po
30	90	4.2	328	1	PXAI_PSEAE	Q915u4 pseudomonas
31	90	4.2	461	1	ATOC_ECOLI	Q06065 escherichia
32	90	4.2	921	1	T2D3_DROME	P47825 drosophila
33	90	4.2	1079	1	CASR_MOUSE	Q9qy96 mus musculu

ALIGNMENTS

RESULT 1

PPA_ECOLI ID_PPA_ECOLI STANDARD; PRT; 432 AA.

AC P07102;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Periplasmic appA protein precursor [Includes: Phosphoanhydride
DE phosphonhydrolase (EC 3.1.3.2) (pH 2.5 acid phosphatase) (AP); 4-
DE ptase (EC 3.1.3.26)]
GN APPA OR B0980.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
[1]
RN SEQUENCE FROM N.A., AND SEQUENCE OF 23-29.
RP STRAIN=K12;
RC MEDLINE=90368616; PubMed=2168385;
RX Dassa J., Marck C., Boquet P.L.;
RA "The complete nucleotide sequence of the Escherichia coli gene appA
RT reveals significant homology between pH 2.5 acid phosphatase and
RT glucose-1-phosphatase."
RL J. Bacteriol. 172:5497-5500 (1990).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=K12 / MG1655;
RC MEDLINE=97426617; PubMed=9278503;
RX Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
RG Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474 (1997).
[3]
RN SEQUENCE FROM N.A.
RP STRAIN=K12;
RC MEDLINE=97061202; PubMed=8905232;
RX Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RM Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RS Sanpei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RY Yano M., Horiuchi T.;
RT "A 718-Kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map."
RL DNA Res. 3:137-155 (1996).
[4]
RN SEQUENCE OF 1-112 FROM N.A.
RP MEDLINE=87271766; PubMed=3038201;
RX Touati E., Danchin A.;
RT "The structure of the promoter and amino terminal region of the pH
RT 2.5 acid phosphatase structural gene (appA) of E. coli: a negative
RT control of transcription mediated by cyclic AMP."
RL Biochimie 69:215-221 (1987).
[5]

RP SEQUENCE OF 1-17 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=92049231; PubMed=1658595;
RA Dassa J., Feilhi H., Marck C., Dion M., Kieffer-Bontemps M.,
RA Boquet P.L.;
RT "A new oxygen-regulated operon in *Escherichia coli* comprises the
RT genes for a putative third cytochrome oxidase and for pH 2.5 acid
RT phosphatase (appa).";
RL Mol. Gen. Genet. 229:341-352(1991).
[6]
RP CHARACTERIZATION, AND SEQUENCE OF 23-34.
RA Greiner R., Jany K.D.;
RL "Characterization of a phytase from *Escherichia coli*.";
RL Biol. Chem. Hoppe-Seyler 372:664-665(1991).
[7]
RP CHARACTERIZATION, AND SEQUENCE OF 23-35.
RX MEDLINE=93256556; PubMed=8387749;
RA Greiner R., Konietzny U., Jany K.D.;
RT "Purification and characterization of two phytases from *Escherichia coli*.";
RL Arch. Biochem. Biophys. 303:107-113(1993).
[8]
RP CHARACTERIZATION.
RX MEDLINE=20161462; PubMed=10696472;
RA Golovan S., Wang G., Zhang J., Forsberg C.W.;
RA van Etten R.L.;
RT "Overexpression, site-directed mutagenesis, and mechanism of
RT *Escherichia coli* acid phosphatase.";
RL J. Biol. Chem. 267:22830-22836(1992).
[9]
RP MUTAGENESIS.
RX MEDLINE=93054596; PubMed=1429631;
RA Ostanin K., Harms E.H., Stevis P.E., Kuciel R., Zhou M.-M.,
RA van Etten R.L.;
RT "Overexpression, site-directed mutagenesis, and mechanism of
RT *Escherichia coli* acid phosphatase.";
RL J. Biol. Chem. 267:22830-22836(1992).
[10]
RP X-RAY CRYSTALLOGRAPHY (2.05 ANGSTROMS).
RX MEDLINE=20122624; PubMed=10655611;
RA Lim D., Golovan S., Forsberg C.W., Jia Z.;
RT "Crystal structures of *Escherichia coli* phytase and its complex with
RT pyruvate".
RL Nat. Struct. Biol. 7:108-113(2000).
CC -1- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O = an
CC alcohol + phosphate.
CC -1- CATALYTIC ACTIVITY: Nyo-inositol hexakisphosphate + H(2)O = 1D-
CC myo-inositol 1,2,3,4,5-pentakisphosphate + phosphate.
CC -1- SUBUNIT: Monomer.
CC -1- SUBCELLULAR LOCATION: Periplasmic.
CC -1- INDUCTION: IN ADDITION TO CAMP-MEDIATED CONTROL, THIS ENZYME IS
CC INDUCED WHEN BACTERIAL CULTURES REACH STATIONARY PHASE. ITS
CC SYNTHESIS IS TRIGGERED BY PHOSPHATE STARVATION OR A SHIFT FROM
CC AEROBIC TO ANAEROBIC CONDITIONS.
CC -1- SIMILARITY: Belongs to the histidine acid phosphatase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M58708; AAA72086.1; -;
DR EMBL; AE000200; AAC74065.1; -;
DR EMBL; D90735; BAA35745.1; -;
DR EMBL; X05471; CAA29031.1; -;
DR EMBL; M58708; -; NOT ANNOTATED CDS.
DR EMBL; S63811; AAB20286.1; -;
DR PIR; B36733; B36733.
DR PDB; 1DKL; 03-AUG-00.
DR PDB; 1DKM; 02-AUG-00.

DR PDB; 1DKN; 03-AUG-00.
DR PDB; 1DKO; 03-AUG-00.
DR PDB; 1DKP; 03-AUG-00.
DR PDB; 1DKQ; 03-AUG-00.
DR SWISS-2DPAGE; P07102; COLI.
DR EcoGene; EGI0049; appA.
DR InterPro; IPR000560; HisAc_phosphatase.
DR Pfam; PF00328; acid_phosphat_1.
DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1.
DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2.
KW Hydrolase; Periplasmic; Signal; Multifunctional enzyme; 3D-structure;
KW Complete proteome.
FT SIGNAL 1 22
FT CHAIN 23 432
FT ACT_SITE 39 39
FT ACT_SITE 326 326
FT DISULFID 99 130
FT DISULFID 155 430
FT DISULFID 200 410
FT DISULFID 404 413
FT CONFLICT 51 66
FT CONFLICT 75 76
FT CONFLICT 112 112
FT STRAND 28 38
FT STRAND 42 42
FT HELIX 49 53
FT TURN 54 54
FT TURN 66 67
FT STRAND 69 69
FT HELIX 71 90
FT TURN 91 92
FT TURN 102 103
FT STRAND 105 109
FT HELIX 113 126
FT TURN 128 129
FT STRAND 134 135
FT TURN 140 141
FT HELIX 145 147
FT TURN 149 153
FT HELIX 159 169
FT TURN 170 171
FT HELIX 174 179
FT TURN 180 181
FT HELIX 182 192
FT TURN 193 193
FT HELIX 194 196
FT HELIX 198 201
FT TURN 202 205
FT HELIX 212 215
FT STRAND 220 222
FT STRAND 227 229
FT HELIX 231 248
FT TURN 249 250
FT HELIX 254 257
FT TURN 258 258
FT HELIX 262 279
FT TURN 280 281
FT HELIX 283 289
FT TURN 290 290
FT HELIX 291 301
FT STRAND 307 308
FT HELIX 310 312
FT STRAND 314 315
FT STRAND 319 324
FT HELIX 326 336
FT TURN 337 337
FT STRAND 342 342
FT STRAND 343 343
FT STRAND 345 345
FT TURN 351 352
FT STRAND 354 362
FT TURN 363 366

MODVTPDAWPTWPKL -> NAGCHPRRMANLAGKT (IN
REF. 3).
EL -> DV (IN REF. 4).
D -> S (IN REF. 4).

PERIPLASMIC APPA PROTEIN.
NUCLEOPHILIC ACCEPTOR.
PROTON DONOR.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AE008748; AAL20049.1; --
DR EMBL; U75949; AAC45604.1; --
DR HSP; P07102; 1DXL.
DR StyGene; SG10595; agp.
DR InterPro; IPR000560; HisAc_pheptase.
DR Pfam; PF00328; acid_phosphat; 1.
DR PROSITE; PS00616; HIS ACID PHOSPHAT_1; 1.
DR PROSITE; PS00778; HIS ACID PHOSPHAT_2; 1.
KW Hydrolase; Periplasmic; Signal; Complete proteome.
FT SIGNAL 1 22 BY SIMILARITY.
FT CHAIN 23 413 GLUCOSE-1-PHOSPHATASE.
FT ACT SITE 40 40 NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).
FT ACT SITE 311 311 PROTON DONOR (BY SIMILARITY).
FT CONFLICT 57 57 T -> P (IN REF. 2).
FT CONFLICT 60 63 AMPA -> TCPP (IN REF. 2).
FT CONFLICT 74 74 K -> T (IN REF. 2).
SQ SEQUENCE 413 AA; 45559 MW; E057667C7F8A7244C CRC64;

Query Match 25.5%; Score 548; DB 1; Length 413;
Best Local Similarity 33.8%; Pred. No. 8.1e-37;
Matches 138; Conservative 66; Mismatches 180; Indels 24; Gaps 10;

QY 1 QSEPE-LKLESVVIVSRHGVRAPE-ITATQMDQVTPDAMPVTPVKGELTPRGGLIAYL 58
DB 23 QTTPEGYQIQVLMMSRHLRPLANNVSVLAQSTPNAPAWDVPGGQITTKGVLEVYM 82

QY 59 GHYWRQLVADGLLPKCGCPQSGQVAILADVDERTKGTGEAFAGLAPDCAITVHTQADT 118
DB 83 GHYTRMLVAQGLIPSGECFAPDPTVYVANSIQTAVATAQFTITGAPGCDIPVHHEKM 142

QY 119 SSPDPLFNPLKTVGCOLDNANTDALLERAGSIAFTGHYQTAFAELERVLNFPQSNLC 178
DB 143 GTMDPTFNFTVITDSSAAFRQAVQA-MEKARSOL-----HLDSEYKLLSEIHTYQDSPSC 196

QY 179 LKREKODESCSLTQALPSELKVSADCVSLTGAVSLSMTEIFLLQQAQMP--EPGWGR 236
DB 197 ----KEKHQCSLIDAKDFTSANVQEPGVQGLKGVNSLVDAFTLQYEGFPMDQVAMGG 252

QY 237 ITDSHOWNTLLSLHNAQFDLLQRTPEVARSRATPLDLIKTALTTPHPKQAYGYTLPTS 296
DB 253 IHTDRQKVLKSLKNGYQDSLFTSPTVAENVAAPLVKYIDKVLV-----ADRVSAK-K 304

QY 297 VLFIAGHDTNLNIGGALELN-WTLPGQPDNTPPGGELVFERWRRLSDNSQIQLVSLVFQ 355
DB 305 VTVIVGHDSNIALLTALDFKPYQLHDOYERTPIGGQLVFORHWDGNANRDLMKIEVVYQ 364

QY 356 TLQOMRDKTFLSLNTPPGVVKLTFLAGEERNAGMCSLAGFTQIVNEA 403
DB 365 SARQLRANAEALTKSPAQRVTLKGCPCV-DANGFCFLDKFDNVNNTA 411

RESULT 3
AGP_ECOLI ID AGP_ECOLI STANDARD; PRT; 413 AA.
AC P19926;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glucose-1-phosphatase precursor (EC 3.1.3.10) (GIPase).
GN AGP OR B1002.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]

CC Between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AE008748; AAL20049.1; --
DR EMBL; U75949; AAC45604.1; --
DR HSP; P07102; 1DXL.
DR StyGene; SG10595; agp.
DR InterPro; IPR000560; HisAc_pheptase.
DR Pfam; PF00328; acid_phosphat; 1.
DR PROSITE; PS00616; HIS ACID PHOSPHAT_1; 1.
DR PROSITE; PS00778; HIS ACID PHOSPHAT_2; 1.
KW Hydrolase; Periplasmic; Signal; Complete proteome.
FT SIGNAL 1 22 BY SIMILARITY.
FT CHAIN 23 413 GLUCOSE-1-PHOSPHATASE.
FT ACT SITE 40 40 NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).
FT ACT SITE 311 311 PROTON DONOR (BY SIMILARITY).
FT CONFLICT 57 57 T -> P (IN REF. 2).
FT CONFLICT 60 63 AMPA -> TCPP (IN REF. 2).
FT CONFLICT 74 74 K -> T (IN REF. 2).
SQ SEQUENCE 413 AA; 45559 MW; E057667C7F8A7244C CRC64;

Query Match 25.5%; Score 548; DB 1; Length 413;
Best Local Similarity 33.8%; Pred. No. 8.1e-37;
Matches 138; Conservative 66; Mismatches 180; Indels 24; Gaps 10;

QY 1 QSEPE-LKLESVVIVSRHGVRAPE-ITATQMDQVTPDAMPVTPVKGELTPRGGLIAYLGH 60
DB 23 QSEPE-LKLESVVIVSRHGVRAPE-ITATQMDQVTPDAMPVTPVKGELTPRGGLIAYLGH 82

QY 61 YWRQLVADGLLPKCGCPQSGQVAILADVDERTKGTGEAFAGLAPDCAITVHTQADTSS 120
DB 83 YWRQLVADGLLPKCGCPQSGQVAILADVDERTKGTGEAFAGLAPDCAITVHTQADTSS 142

QY 121 PDPLFNPLKTVGCOLDNANTDALLERAGSIAFTGHYQTAFAELERVLNFPQSNLC 180
DB 143 PDPLFNPLKTVGCOLDNANTDALLERAGSIAFTGHYQTAFAELERVLNFPQSNLC 202

QY 181 REKODESCSLTQALPSELKVSADCVSLTGAVSLSMTEIFLLQQAQMPGPGWGRITDS 240
DB 203 REKODESCSLTQALPSELKVSADCVSLTGAVSLSMTEIFLLQQAQMPGPGWGRITDS 262

QY 241 HOWNTLISLHNAQFDLLQRTPEVARSRATPLDLIKTALTTPHPKQAYGYTLPTS 300
DB 263 HOWNTLISLHNAQFDLLQRTPEVARSRATPLDLIKTALTTPHPKQAYGYTLPTS 322

QY 301 AGHDTNLNIGGALELN-WTLPGQPDNTPPGGELVFERWRRLSDNSQIQLVSLVFQ 360
DB 323 AGHDTNLNIGGALELN-WTLPGQPDNTPPGGELVFERWRRLSDNSQIQLVSLVFQ 382

QY 361 RDKTFLSLNTPPGVVKLTFLAGEERNAGMCSLAGFTQIVNEAIPACSL 410
DB 383 RDKTFLSLNTPPGVVKLTFLAGEERNAGMCSLAGFTQIVNEAIPACSL 432

RESULT 2
AGP_SALTY ID AGP_SALTY STANDARD; PRT; 413 AA.
AC Q33921;
DT 15-JUL-1998 (Rel. 36, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glucose-1-phosphatase precursor (EC 3.1.3.10) (GIPase).
GN AGP OR STM117.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P., Courtney L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D., Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E., Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M., Waterston R., Wilson R.K.;
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium LT2.";
RT Nature 413:852-856(2001).
RL [2]
RN SEQUENCE OF 1-249 FROM N.A.
RP STRAIN=TN1379;
RX MEDLINE=97405892; PubMed=9260936;
RA Gupta S.D., Wu H.C., Rick P.D.;
RT "A Salmonella typhimurium genetic locus which confers copper tolerance on copper-sensitive mutants of Escherichia coli.";
RL J. Bacteriol. 179:4977-4984(1997).
CC -/- CATALYTIC ACTIVITY: D-glucose 1-phosphate + H(2)O = D-glucose + phosphate.
CC -/- SUBUNIT: Homodimer (By similarity).
CC -/- SUBCELLULAR LOCATION: Periplasmic.
CC -/- SIMILARITY: Belongs to the histidine acid phosphatase family.
CC [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-31.
RC STRAIN=K12;
RX MEDLINE=90130318; PubMed=2153660;
RA Pradel E., Marck C., Boquet P.L.;
RT "Nucleotide sequence and transcriptional analysis of the Escherichia
coli agp gene encoding periplasmic acid glucose-1-phosphatase";
RL J. Bacteriol. 172:802-807(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12";
RL Science 277:1433-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Ohshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
Mori H., Mocomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
Sampel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map";
RL DNA Res. 3:137-155(1996).
RN [4]
RP SEQUENCE OF 23-34.
RC STRAIN=K12 / EMG2;
RX MEDLINE=97443975; PubMed=9298646;
RA Link A.J., Robison K., Church G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
in the genome of Escherichia coli K-12";
RL Electrophoresis 18:1259-1313(1997).
CC -1- FUNCTION: ABSOLUTELY REQUIRED FOR THE GROWTH OF E. COLI IN A HIGH-
PHOSPHATE MEDIUM CONTAINING G-1-P AS THE SOLE CARBON SOURCE.
CC -1- CATALYTIC ACTIVITY: D-glucose 1-phosphate + H(2)O = D-glucose +
phosphate.
CC -1- ENZYME REGULATION: INDEPENDENT FROM INORGANIC PHOSPHATE
AVAILABILITY, AND APPARENTLY SUBMITTED TO CATABOLITE REPRESSION,
IT IS POSITIVELY CONTROLLED BY CAMP AND THE CAMP RECEPTOR PROTEIN.
CC -1- SUBUNIT: Homodimer.
CC -1- SUBCELLULAR LOCATION: Periplasmic.
CC -1- MISCELLANEOUS: OPTIMALLY ACTIVE AROUND PH 4.
CC -1- SIMILARITY: Belongs to the histidine acid phosphatase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; M33807; AAA23426.1; -.
DR EMBL; AE000202; AAC74087.1; -.
DR EMBL; D30737; BAA35769.1; -.
DR EMBL; D50738; BAA35779.1; -.
DR PIR; JVO087; JVO087.
DR HSSP; P07102; 1DKM.
DR EcoGene; EG10033; agp.
DR InterPro; IPR000560; HisAc_phsphtse.
DR Pfam; PF00328; acid_phosphat; 1.
DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
KW Hydrolase; Periplasmic; Signal; Complete proteome.
FT SIGNAL 1 22
FT CHAIN 23 413 GLUCOSE-1-PHOSPHATASE.
FT ACT_SITE 40 40 NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).

FT ACT_SITE 311 311 PROTON DONOR (BY SIMILARITY).
SQ SEQUENCE 413 AA; 45683 MW; ADADAD3639D0D6AB CRC64;

Query Match 25.3%; Score 544; DB 1; Length 413;
Best Local Similarity 33.1%; Pred. No. 1.7e-36;
Matches 137; Conservative 71; Mismatches 170; Indels 36; Gaps 12;

QY 1 QSEPE-LKLESVWIVSRHGVRAP-TKATQLMQDVTDPDAPFTWPKVKGELTPRGELIAYL 58
DB 23 QTVEGYQLQVLMWSRHLRAPLANNQSVLEQSTPNKPEWDVPGGQLTKGGVLEVM 82

QY 59 GHYWRQLVADGLLPKCGCQSQGOVIAIADVDERTKRTGEAFAPAGLAPDCAITVHTQADT 118
DB 83 GHYVREWLAEQGMVKGSCPPPTVYAYANSLQRTVATRAQFFITGAPFGCDIPVHHQEKM 142

QY 119 SSPDPLFNPLKTVGCQLDNANVTD---AILERAGGSIADFTGHYQTAFRELRLVNFPOS 175
DB 143 GTMDPTFNPVTD-----DSAAFSEQAVAMEKLSKL-QLTDSVQL-----LEKIVNPKDS 193

QY 176 NLCLKREKQDESCSLTQALPSELKVSADCVSLTQAVSLASMLTEIFILQQAQGNP--EPG 233
DB 194 PAC----KEQCCSLVDGKNTFSAKYQQEFGVSGFLKVGNSLVDAFTLQYEGFPMQOVA 249

QY 234 WGRITDSHOWNTLLSLHNAQFDLLQRTPEVARSRATPLLDLIKLTALTTPHPPQKQAYVTL 293
DB 250 WGEIKSDQWKVSLKXGYQDSLFTSFVARNVAKPLVSYIDKAL-----VTD 298

QY 294 PTS---VLFIAGHDTNLANIGALELN-WTLPGQPDNTPPGELVFFRWRRLSDNSQWIQ 349
DB 299 RTSAPKITVLVGHDSNIAISLLTALDFKPYQLHDQNETPIGKIVFORWHDSEKANRDLMK 358

QY 350 VSLVFPQTLQOMRDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGTQIVNEA 403
DB 359 IEYVQSAEQURNADALTQAPORVILELGGC-PIDADGFCPMKDFSDVINEA 411

RESULT 4
ID AGP_PROBE STANDARD; PRT; 417 AA.
AC Q52309.
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Glucose-1-phosphatase precursor (EC 3.1.3.10) (Gipase).
GN AGP.
OS Providencia rettgeri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Providencia.
OX NCBI_TaxID=587;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PV7;
RA Riccio M.L., Chiesurin A., Lombardi G., Satta G.;
RL Submitted (SEP-1992) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: D-glucose 1-phosphate + H(2)O = D-glucose +
phosphate.
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Periplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the histidine acid phosphatase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; X68201; CAA48288.1; -.
DR PIR; S25627; S25627.
DR HSSP; P07102; 1DKL.
DR InterPro; IPR000560; HisAc_phsphtse.
DR Pfam; PF00328; acid_phosphat; 1.

FT CARBOHYD 267 267 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 29 29 Q -> R (IN REF. 1 AND 2).
SQ SEQUENCE 423 AA; 48316 MW; 1A2CBFA40ED3724B CRC64;
Query Match 6.1%; Score 132; DB 1; Length 423;
Best Local Similarity 23.3%; Pred. No. 0.0037;
Matches 97; Conservative 43; Mismatches 173; Indels 104; Gaps 20;
Qy 8 LESVIVSRHGVRAPTKATQMDVTPDAWTPWVKLGELTPRGGLIAYLGHYWRQLV 67
Db 33 LRFVTLRYHGRSPVKT---YKDPYQBEWPGQFGQLTKGMLQHWELGQALRQY- 87
Qy 68 ADGLLPKCGCQSQGQVAILADV-ERTKTEGFAAGLAPDCAITVHTQADTSSPDPL- 124
Db 88 -HGFL---NTSYHRQEVVVRSTDFRTLMSAEANLAGLFP-----PQMGQR 129
Qy 125 FNP-LKTVGCOLDNANVTDAILERAGGSIADFTGHYQTAFLERVLNFPQSNLCLKREK 183
Db 130 FNPNISWQPIVHTVPIE-----DRLLKFPGLGPCRYEQL 165
Qy 184 ODESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFL-----LQAQGMPEP 232
Db 166 QNE-----TRQTFEQNENSSRNAQFLDMVANETGLDUTLETWNVYDTLFCETHGLRUP 221
Qy 233 GWGRITDSHOWNTLLSLHNAQFDLL---QRTPEVARSRATPLDLINKALTALTPHPKOAY 289
Db 222 PW---ASPQTMORLSRLKDFSRFLFGIYQAEKARLQGGVLLAQRKNTLTWATTSQ-- 276
Qy 290 GVTLPTSVLFTAGHDTNLANLGGALELNWTLPGQDNTPPGGLVPERWRRLSDNSQWIQ 349
Db 277 ---LP-KLLVYSADHTTLVALQMALDVI-----NGEQAPYASCHIFELYQEDSGN---FS 324
Qy 350 VSLVFQTLQOMRDKTPLSLNTPPGEVKTLGACBNAQMGCSLAGTQTIVNEARIP 406
Db 325 VEMTF---RNESKAPWPLSLP-----GCPRH-----CPLQDFLRL-TEPVVP 363
RESULT 6
FPAL RAT
ID FPAL RAT STANDARD; PRT; 423 AA.
AC P20611;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Lysosomal acid phosphatase precursor (EC 3.1.1.3.2) (LAP).
GN ACP2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89350910; PubMed=2764916;
RA Hienno M., Fujita H., Noguchi Y., Kono A., Kato K.;
RT "Isolation and sequencing of a cDNA clone encoding acid phosphatase
in rat liver lysosomes";
RL Biochem. Biophys. Res. Commun. 162:1044-1053(1989).
CC -!- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O = an
CC alcohol + phosphate.
CC -!- SUBCELLULAR LOCATION: Lysosomal.
CC -!- SIMILARITY: Belongs to the histidine acid phosphatase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M27893; AAA40744.1; -

DR A33395; A33395.
DR HSP; P20446; IRPA.
DR InterPro; IPR000560; HisAc_phosphatase.
DR Pfam; PF0328; acid_phosphat; 1.
DR PROSITE; PS00616; HIS_ACID_PHOSPHAT 1; 1.
DR PROSITE; PS00778; HIS_ACID_PHOSPHAT 2; 1.
KW Hydrolase; Signal; Glycoprotein; Lysosome.
FT SIGNAL 1 30
FT CHAIN 31 423 LYSOSOMAL ACID PHOSPHATASE.
FT DISULFID 159 370 BY SIMILARITY.
FT DISULFID 212 310 BY SIMILARITY.
FT DISULFID 345 349 BY SIMILARITY.
FT ACT_SITE 42 42 NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).
FT ACT_SITE 84 84 BY SIMILARITY.
FT ACT_SITE 286 286 PROTON DONOR (BY SIMILARITY).
FT CARBOHYD 92 92 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 167 167 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 177 177 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 197 197 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 267 267 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 423 AA; 48319 MW; 6F4C4819CADD4496 CRC64;
Query Match 5.3%; Score 114; DB 1; Length 423;
Best Local Similarity 22.7%; Pred. No. 0.11;
Matches 95; Conservative 42; Mismatches 175; Indels 106; Gaps 20;
Qy 8 LESVIVSRHGVRAPTKATQMDVTPDAWTPWVKLGELTPRGGLIAYLGHYWRQLV 67
Db 33 LRFVTLRYHGRSPVKA---YKDPYQBEKVGQFGQLTKGMLQHWELGQALRQY- 87
Qy 68 ADGLLPKCGCQSQGQVAILADV-ERTKTEGFAAGLAPDCAITVHTQADTSSPDPLFN 126
Db 88 -HGFL---NASYHRQEVVVRSTDFRTLMSAEANLAGLFP-----TEVQHFNISWQ 137
Qy 127 PLKTVGCOLDNANVTDAILERAGGSIADFTGHYQTAFLERVLNFPQSNLCLKREKQDE 186
Db 138 PT-----EVHTVPIE-----DRLLKFPGLGPCRYEQLQNE 168
Qy 187 SCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFL-----LQAQGMPEPGWG 235
Db 169 ---TRQTFEQNENSSRNAQFLDMVANETGLMNLTLTETWVYDTLFCETHGLLPPW- 223
Qy 236 RITDSHOWNTLLSLHNAQF-----DLLQRTPEVARSRATPLDLINKALTALTPHPKOQA 288
Db 224 ---ASPQTMORLSRLKDFSRFLFGIYQAEKARLQGGVLLAQRKNTLTWATTSQF 277
Qy 289 YGVTLPTSVLFTAGHDTNLANLGGALELNWTLPGQDNTPPGGLVPERWRRLSDNSQWI 348
Db 278 ---PKLLVYSADHTTLVALQMALNVI-----NGKQAPYASCHIFELYQ--EDNGNF- 323
Qy 349 QVSLVFQTLQOMRDKTPLSLNTPPGEVKTLGACBNAQMGCSLAGTQTIVNEARIP 406
Db 324 SVEMTFRNDSK---KAPW-----PLTLPGCPHR-----CPLQDFLRL-TEPVIP 363
RESULT 7
PAX8_MOUSE
ID PAX8_MOUSE STANDARD; PRT; 457 AA.
AC Q00288; P97342;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Paired box protein Pax-8.
GN PAX8 OR PAX-8.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]


```

DR ENBL; L02420; AAI16897.1; -.
DR PIR; JN0890; JN0890.
DR PDB; 1QFX; 19-APR-00.
DR InterPro; IPR000560; HisAc_phosphatase.
DR Pfam; PF00328; acid_phosphat_1; 1.
DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
KW Hydrolase; Glycoprotein; Signal; 3D-structure.
FT SIGNAL 1 19
FT CHAIN 20 479
FT ACT SITE 82 82
FT ACT SITE 337 337
FT ACT SITE 71 387
FT DISULFID 128 472
FT DISULFID 216 441
FT DISULFID 225 298
FT DISULFID 413 421
FT CARBOHYD 191 191
FT CARBOHYD 315 315
FT CARBOHYD 458 458
FT HELIX 41 43
FT HELIX 45 47
FT TURN 50 51
FT TURN 69 70
FT STRAND 71 81
FT STRAND 85 85
FT HELIX 89 103
FT HELIX 112 119
FT HELIX 127 129
FT TURN 130 131
FT STRAND 133 133
FT TURN 138 139
FT HELIX 141 156
FT HELIX 157 159
FT STRAND 166 171
FT STRAND 173 173
FT HELIX 174 188
FT HELIX 189 191
FT HELIX 192 195
FT STRAND 196 201
FT HELIX 205 207
FT TURN 208 209
FT TURN 224 227
FT TURN 233 233
FT HELIX 234 245
FT TURN 247 248
FT HELIX 253 266
FT TURN 267 269
FT TURN 274 275
FT HELIX 276 278
FT TURN 281 298
FT TURN 301 302
FT TURN 304 305
FT HELIX 306 323
FT HELIX 325 328
FT STRAND 331 336
FT HELIX 338 348
FT TURN 349 349
FT TURN 364 365
FT HELIX 370 372
FT TURN 373 373
FT TURN 376 377
FT STRAND 379 387
FT STRAND 390 390
FT TURN 391 392
FT STRAND 393 393
FT STRAND 396 403
FT TURN 404 405
FT STRAND 406 408
FT STRAND 410 410
FT TURN 411 412
FT STRAND 413 413
FT HELIX 417 419

Query Match
5.1%; Score 109.5; DB 1; Length 479;
Best Local Similarity 19.2%; Pred. No. 0.29;
Matches 91; Conservative
QY 7 KLESVIVSRHGVRAPTKATQLMQDVTTPDAWTPWPKL-----GELT----- 48
Db 72 EVDQVIMVKRHGRYPSPSAG--KDIEEALAKYISINTTEYKGDGLAFPLNDWTYYVNECY 129
QY 49 -----PRGGELIAY-----LGHYWRQRLVADGLLPKCGCPQSQGVAIADVDE 91
Db 130 YNAETTSGPYAGLLDAYNHGNDYKARYGHLWNGETVVPFF-----SSGYGRVI---E 178
QY 92 RTRKTGEA-FAAGLAPDCAITVHTQADTSSPDPLFNPLKTVGQOLDNAVNTDAILERAGG 150
Db 179 TARXFGEGFFGYNYSTNAALNIISESEVMGADSL-----TPTCDTNDQDTCDNLT--- 229
QY 151 SIADFTGHYOTAFRELERVLNFPQSINLCKR-EKQDESCSLTQALPSELKVSADCVSLTG 209
Db 230 -----YQ-----LPQFVAAALNSQNGMNL-----ASDVYNL-- 259
QY 210 AVSLASMLTEIFLLQOAGMPEPGWGRITDSHOWNTL-----LSLH----- 250
Db 260 -MVNASF-----ELNARFFSNWINAFTQDENVSFGYVEDLNYYXCAGFGDKNMAAVG 310
QY 251 ----NAOFDILLQRTPEVARSRATPLLDLIKTALTPHPQKQAYGVTLPTSVLFTIAGHDTN 306
Db 311 AVYANASLTLLNQPKEA-----GSLFPNFAHDTN 340
QY 307 LANLGGALEL---NWTLP-----GQP-----DNTPPGSELVFERM-----RRLSDNSQW 349
Db 341 ITPILAAALGVLPNEDPLDRVAFGNFYSIGNIVPMGGHLTIERLSQATALSDEGTYYR 400
QY 350 VSLVFQTLQMRDKTPLSLNTPPEGVKLTLAGCEERNAQGMCSLAGTQIVNE 402
Db 401 LVL-----NEAVLPFNDCVSGPGYS-CPLANTISLKNK 432

RESULT 9
RY44 DROME
ID RY44 DROME STANDARD; PRT: 5127 AA.
AC Q24498; Q24321; Q24499; Q24500; Q24501; Q8MKS3; Q8MKS4; Q8MKS5;
AC Q9V4Y7;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ryanodine receptor 44F.
GN RYA-R44F OR DEY OR RYR OR CG10844.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS A; B; C AND D), FUNCTION, TISSUE
RP SPECIFICITY, AND DEVELOPMENTAL STAGE.
RX MEDLINE=94102409; PubMed=8276118;
RA Takeshima H., Nishi M., Iwabe N., Miyata T., Hosoya T., Masai I.,
RA Hotta Y.;
RT Isolation and characterization of a gene for a ryanodine
RT receptor/calcium release channel in Drosophila melanogaster.;
RL FEBS Lett. 337:81-87(1994).
RN [2]

```

RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RX MEDLINE=2019606; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Gallie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhendari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fierro C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinart K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.B., Wang X.,
RA Swirskas R., Tector C., Turner E., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of *Drosophila melanogaster*."
RT Science 287:2185-2195(2000).
[3]
RP REVISIONS, AND ALTERNATIVE SPLICING.
RX MEDLINE=2242606; PubMed=12537572;
RA Miera S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaninker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Ceiniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a
RT systematic review".
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
[4]
RP SEQUENCE OF 4619-5127 FROM N.A., DEVELOPMENTAL STAGE, AND TISSUE
SPECIFICITY.
RX TISSUE=Brain;
RX MEDLINE=93202018; PubMed=1338312;
RA Haan G., Roobash M.;
RT "Drosophila homologs of two mammalian intracellular Ca(2+)-release
RT channels: identification and expression patterns of the inositol
RT 1,4,5-trisphosphate and the ryanodine receptor genes".
RL Development 116:967-975(1992).
[5]
RP FUNCTION, AND TISSUE SPECIFICITY.
RX MEDLINE=20283930; PubMed=10811919;
RA Sullivan K.M.C., Scott K., Zuker C.S., Rubin G.M.;
RT "The ryanodine receptor is essential for larval development in
RT *Drosophila melanogaster*".
RL Proc. Natl. Acad. Sci. U.S.A. 97:5942-5947(2000).
CC -!- FUNCTION: Communication between transverse-tubules and
CC sarcoplasmic reticulum. Contraction of muscle is triggered by
CC release of calcium ions from SR following depolarization of T-
CC tubules (By similarity).
CC -!- FUNCTION: Intracellular calcium channel that is required for
CC proper muscle function during embryonic development and may be
CC essential for excitation-contraction coupling in larval body wall
CC muscles.
CC -!- SUBUNIT: Homotetramer (Probable).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Name=A;
CC IsoId=Q24498-2; Sequence=Displayed;
CC Name=B;
CC IsoId=Q24498-3; Sequence=VSP_050196;
CC Name=C;
CC IsoId=Q24498-4; Sequence=VSP_050195;
CC Name=D;
CC IsoId=Q24498-5; Sequence=VSP_050195, VSP_050196;
CC TISSUE SPECIFICITY: During embryonic stages 9-10, expression is
CC seen in mesoderm of all segments in progenitors of the cephalic
CC and somatic muscles. Adults exhibit high expression in tubular
CC 'jump' muscles of thorax and leg, and lower expression in the
CC brain, ventral ganglion, head muscles and proboscis muscles.
CC -!- DEVELOPMENTAL STAGE: Abundant in 6-12 hour embryos, reduced
CC expression in second and third instar larval stages and adults.
CC -!- MISCELLANEOUS: The calcium release channel is modulated by calcium
CC ions, magnesium ions, ATP and calmodulin (By similarity).
CC -!- MISCELLANEOUS: The calcium release channel activity resides in the
CC C-terminal region while the remaining part of the protein
CC constitutes the 'foot' structure spanning the junctional gap
CC between the SR and the T-tubule. It is possible that the foot
CC structure interacts with the cytoplasmic region of the
CC dihydropyridine receptor (By similarity).
CC -!- MISCELLANEOUS: Ryanodine is an alkaloid that binds to the Ca-
CC release channel in junctional SR and modulates its activity (By
CC similarity).
CC -!- SIMILARITY: Belongs to the ryanodine receptor family.
CC -!- SIMILARITY: Contains 5 MIR domains.
CC -!- SIMILARITY: Contains 3 SPRY domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; D17389; BAA04212.1; -;
CC EMBL; D17389; BAA41469.1; -;
CC EMBL; D17389; BAA41470.1; -;
CC EMBL; D17389; BAA41471.1; -;
CC EMBL; AE003835; AAF59036.2; -;
CC EMBL; AE003835; AAM71082.1; -;
CC EMBL; AE003835; AAM71083.1; -;
CC EMBL; AE003835; AAM71084.1; -;
CC EMBL; Z18536; CAA79221.1; -;
CC FlyBase; FBgn0011286; Rya-r44f.
CC GO; GO:0016021; C:integral to membrane; NAS.
CC GO; GO:0005219; P:ryanodine-sensitive calcium-release channel. . .; TAS.
CC GO; GO:0006816; P:calcium ion transport; TAS.
CC GO; GO:0006936; P:muscle contraction; IMP.
CC InterPro; IPR000699; Ca-rel_channel.
CC InterPro; IPR001682; Ca/Na_Pore.
CC InterPro; IPR002048; EF-hand.
CC InterPro; IPR005821; Ion_trans.
CC InterPro; IPR003608; MIR.
CC InterPro; IPR001215; Ryanodn_receptor.
CC InterPro; IPR003032; RYR.
CC InterPro; IPR003877; SPRY_receptor.
CC Pfam; PF00520; Ion_trans; 1.
CC Pfam; PF02815; MIR; 4.
CC Pfam; PF01365; RYDR_ITPR; 2.

DR Pfam; PF002026; RYR; 4.
DR Pfam; PF00622; SPRY; 3.
DR PRINTS; PR00795; RYANODINER.
DR SMART; SM00472; MIR; 4.
DR SMART; SM00449; SPRY; 3.
DR PROSITE; PS00919; MIR; 5.
KW Repeat; Transmembrane; Receptor; Ionic channel; Calcium channel;
KW Repeat; Transmembrane; Glycoprotein; Phosphorylation;
KW Alternative splicing.
FT DOMAIN 1 3232
FT TRANSNEM 3233 3253
FT TRANSNEM 3301 3321
FT TRANSNEM 3376 3396
FT TRANSNEM 3963 3983
FT TRANSNEM 4452 4472
FT TRANSNEM 4651 4671
FT TRANSNEM 4683 4703
FT TRANSNEM 4742 4762
FT TRANSNEM 4885 4905
FT TRANSNEM 4924 4944
FT TRANSNEM 4963 4983
FT TRANSNEM 5004 5024
FT DOMAIN 94 148
FT DOMAIN 155 200
FT DOMAIN 210 264
FT DOMAIN 270 328
FT DOMAIN 336 394
FT DOMAIN 663 802
FT DOMAIN 1091 1216
FT DOMAIN 1529 1672
FT DOMAIN 4686 4711

Query Match 5.18; Score 109; DB 1; Length 5127;
Best Local Similarity 20.98; Pred. No. 7.8;
Matches 89; Conservative 45; Mismatches 133; Indels 158; Gaps 18;
QY 42 VKLGELTPRGEL---IAYIGHYWRQLVA-----DGLPKGCGPQSGQVAIIADVDEIT 93
DB 3934 LNLGIALIRGENDIQMGLNHLKXKDVGFPSIAGLWNSCS-----VLDLDAPE 3984
QY 94 RKTGEAPAGLAPDCAITVHTQADTSSPDLF-----NPLKTGCQCLDNA 138
DB 3985 RNT-KASGLGVSGAAGERKMDAETCALRFIQTCEGHNLEWQNYLRTQAGNTTV 4043
QY 139 NV---TDAILERAGGSIAFTGHYQTAFARELVNFPQSNLCLKREKODESCSLTQALP 195
DB 4044 NVVICTVYLLRLQESIMDFWYHSS-----LQQAQGWPEPGWGHIITDSHOWNTL 4069
QY 196 SELKVSADCVSLTGAVSLAM-----LLEIF-----LLQQAQGWPEPGWGHIITDSHOWNTL 246
DB 4070 XEIIDPAGKANFFKAIIVASQVNTLTVIQGPCTLNQQA-----LAHSLRWDAV 4119
QY 247 -----LSLENAQFDLLQRTPEVARSRAATPLDLIKTALTTPHPQKQAYGVTL 293
DB 4120 GGFLFLFSHMQDKLSKSSQVDLLKE-----LLNLQKDMIT----- 4155
QY 294 PTVSLFAGHDTNLANLGGALELNWTLPGQDNTTPGGELV---FEWRRLSD----- 343
DB 4156 ---MLSLMELGNVNGTIGKQVDDTLVESAENV-----ELIKYDFMFLKADLIESPSF 4207
QY 344 -----NSQWQVSLVFTQLQWMDKTPSLNTPPGEVKLTLAGCERNAQGCMSLAGFT 397
DB 4208 HEVDKKEGV-----TPKDFREKMEQSKNYTPEWDFILACC-ERNHEGKIDYRAFV 4259
QY 398 QIVNE 402
DB 4260 EHFHE 4264

RESULT 10
PHYB_ASPNG
ID PHYB_ASPNG
AC P34754;

DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 3-phytase B precursor (EC 3.1.3.8) (Myo-inositol-hexaphosphate
DE 3-phosphohydrolase B) (3 phytase B) (Myo-inositol hexakisphosphate
DE 3-phosphohydrolase B).
GN PHYB.
OS Aspergillus niger.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eutotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OC NCBI_TaxID=5061;
OX (1)
RP SEQUENCE FROM N.A., AND SEQUENCE OF 20-101; 133-141 AND 376-399.
RX MEDLINE=93371452; PubMed=7916610;
RA Ehrlich K.C., Montalbano B.G., Mullaney E.J., Dischinger H.C. Jr.,
RA Ullah A.H.J.;
RT "Identification and cloning of a second phytase gene (phyB) from
RT Aspergillus niger (ficcum).";
RL Biochem. Biophys. Res. Commun. 195:53-57(1993).
CC -!- FUNCTION: Catalyzes the hydrolysis of inorganic orthophosphate
CC from phytate.
CC -!- CATALYTIC ACTIVITY: Myo-inositol hexakisphosphate + H2O = 1D-
CC myo-inositol 1,2,4,5,6-pentakisphosphate + phosphate.
CC -!- SIMILARITY: Belongs to the histidine acid phosphatase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L20567; AAA02934.1; -
DR HSP; P34755; IOPX.
DR InterPro; IPR000560; HisAc_phosphatase.
DR Pfam; PF00328; acid_phosphat; 1.
DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
KW Hydrolase; Glycoprotein; Signal.
FT SIGNAL 1 19
FT CHAIN 20 479
FT ACT_SITE 81 81
FT ACT_SITE 82 82
FT ACT_SITE 337 337
FT CARBOHYD 106 106
FT CARBOHYD 191 191
FT CARBOHYD 227 227
FT CARBOHYD 250 250
FT CARBOHYD 315 315
FT CARBOHYD 425 425
FT CARBOHYD 442 442
FT CARBOHYD 458 458
FT SEQUENCE 479 AA; 52611 MW; 395D4DA2B50PDFC4 CRC64;
Query Match 5.0%; Score 108.5; DB 1; Length 479;
Best Local Similarity 20.2%; Pred. No. 0.35;
Matches 95; Conservative 44; Mismatches 147; Indels 185; Gaps 25;
QY 7 KLESVIVSRHGRVAPT-----KATQMODVT-PDAW----PTWPVKL 44
DB 72 EVDQVIMVKRHEERYSPSAGKSTEEALAKVYSINTTEYKGLDPLNDWTYVVPNECYN 131
QY 45 GELT--PRGELIAY-----LGHYWRQLVADGLLPKCGCPQSQVAILADVDEIT 93
DB 132 AETTSQPYAGLLDAYNHGNDYKARYGHLWNGETVVVFF-----SSGYGVI-----ETA 180
QY 94 RKTGEA-FAAGLAPDCAITVHTQADTSSPDLNPLNPLKTVGQCLDNANVTDAILERAGGS 152
DB 181 RKFGSGFGYNYSTNAALNIISEVWGADSL-----TPTCDTNDQTTCDNLT----- 229
QY 153 ADFTGHYQTAFARELVNFPQSNLCLKREKODESCSLTQALPSELKVSADCVSLTGA 211

```
Db 230 -----YQ-----LPQKVAARLNSQNGHNTL-----ASDVNLT-----I 260
QY 212 SLASMLTEIFLLQQAQMPFGNGRITDTHSHQNTL-----LSLH----- 250
Db 261 VWASF-----ELNARPFNSWNTAFQDEWVSFGVVEDLNYYCAGPGDKNMAAVGAV 312
QY 251 --NAQFDLQRTPEVARSRAPLDLIKALTATPHPPQKQAVGVTLPVSLFIAGHDTNLA 308
Db 313 YANASLTLLNGPK-----EAGPL-----FFNFAHDTNIT 342
QY 309 NLGGALEL-----NWTL-----DNTPPGGELVFERW-----RRLSNDSNQIOWS 351
Db 343 PILAALGVLIENEDLFLDRVAFNGPYSGNIVPMGGHLLTIERLSQATALSXGTYVRLV 402
QY 352 LVFQTLQWRDKTPLSLNTPGVEVKLILAGCEERNAQMGSLAGFTQIVNE 402
Db 403 L-----NEAVLPNDCTSGFGYS-CPLANYTSILNK 432

RESULT 11
PAX8_CANFA
ID PAX8_CANFA STANDARD; PRT; 459 AA.
AC P47240; P47241;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Paired box protein Pax-8.
GN PAX8 OR PAX-8.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 8A AND 8G).
RC TISSUE=Thyroid;
RX MEDLINE=96254068; PubMed=8652674;
RA van Renterghem P.H.G., Vassart G., Christophe D.;
RT "Pax 8 expression in primary cultured dog thymocyte is increased by cyclic AMP.";
RL Biochim. Biophys. Acta 1307:97-103(1996).
CC -!- FUNCTION: THOUGHT TO ENCODE A TRANSCRIPTION FACTOR. IT MAY HAVE A ROLE IN KIDNEY CELL DIFFERENTIATION. MAY PLAY A REGULATORY ROLE IN MAMMALIAN DEVELOPMENT.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=8A;
CC IsoId=P47240-1; Sequence=Displayed;
CC Name=8G;
CC IsoId=P47240-2; Sequence=VSP 002371;
CC -!- SIMILARITY: Contains 1 paired box domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X83591; CAA58571.1; -
DR EMBL; X83592; CAA58572.1; -
DR PIR; S70361; S52250.
DR PIR; S70362; S52251.
DR HSSP; P26367; 6PAX.
DR TRANSFAC; T02930; -.
DR InterPro; IPR001523; Paired_box.
DR Pfam; PF00292; PAX; 1.
DR PRINTS; PR00027; PAIREDBOX.
DR SMART; SM00351; PAX; 1.
DR PROSITE; PS00034; PAIRED_BOX; 1.
```

```
KW Paired box; Developmental protein; Nuclear protein; DNA-binding;
KW Transcription regulation; Differentiation; Alternative splicing.
FT DOMAIN 9 133
FT VARSPLIC 302 459
FT -----
FT DPSPFAIKQETPEVSSSSSTPSSLSAFLDLQVGGAGV
FT PAGAASVPFNAPFASVYQFTGQALLSGREMGVTPPGY
FT PPHIPTSGGVSASAIAGVAGSEYSGNAVGHYTSYSGE
FT AWRFPNSLLSPYVYSTSRPSAPPTATAFDHL -> AP
FT FEWICKSNAQGSQVPSRSPMPMLPCTGSSRRAPSS
FT QSERWDEHCDTHTPPVDRAANPLLPQAWQVPHIT
FT VPHQGVRRHPPPRPLTTCSCYDGSN (in isoform 8G).
FT -----
FT FTID=VSP 002371.
SQ SEQUENCE 459 AA; 48839 MW; 4CC00164A8E56935 CRC64;

Query Match 5.0%; Score 107.5; DB 1; Length 459;
Best Local Similarity 18.7%; Pred.No.0.4; 130; Indels 171; Gaps 16;
Matches 81; Conservative 51; Mismatch 130; Indels 171; Gaps 16;

QY 4 PELKESVIVSRHGVRAPTKATQL-----MODVTPDAWTPVVKLGL-----TP 49
Db 27 PEVVRQRIVDLAHQGVPCDISRQLRVSHGCVSKILGRYETGSRPGVIGGSKPKVATP 86
QY 50 RGGELIAYLGHYWRQ-----RLVADGL-----LP 73
Db 87 K--VVEKIGDYKRONPTMFAMEIRDLAELGVCNDTVPSVSSINRIKTKVQGFNLP 143
QY 74 KGCQ-----POS-----GQVATIAVDERTKRTGEAF 100
Db 144 MDSCVATKSLSPGHTLIPSSAVTPPESQSDSLGSTYSINGLLGIAQPGSDSKRWDDSD 203
QY 101 AAGLAPDCAITVHTQADTSSDPPLFNPLKTGVQCLDNANVTAILERAGGSTADFTGHVQ 160
Db 204 ----QDSCRLSIDSSSSGPR---KHLRTDAFSQHLEPLECPFERQ-----HYP 247
QY 161 TAPRELERYLNFQSNLCLKREKQDESCLTQALPSLKVSAQCVSL-----TGAV 211
Db 248 EAY-----ASPSHTKGEQGLYPLPLNSAALDDGKA 278
QY 212 SLASMLTEIFLLQQAQMPFGNGRITDTHSHQNTLLSLHNAQDFLLQRTPEVARSRATP- 270
Db 279 TLTPSNTP-----GRNLSTHQTYPVAVDPHSPFAIKQETPEVSSSSSTPS 324
QY 271 -----LLDLIKALTAPHPQKQAVGVTLPVSLF-----IAGHDTNLANLGALELNT 319
Db 325 SLSSSAFLDLQVGGAGVPA-----GASVPPNAPFASVYQFTGQALLSGREMGVGT 379
QY 320 LFGQPDNTPPGGE 332
Db 380 LFGYPPHIPTSGQ 392

RESULT 12
PAX8_HUMAN
ID PAX8_HUMAN STANDARD; PRT; 450 AA.
AC Q06710; Q09155; Q16337; Q16338; Q16339;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Paired box protein Pax-8.
GN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND VARIANT LEU-329.
RC TISSUE=Kidney;
RX MEDLINE=93170167; PubMed=1337742;
RA Poleev A., Fickerscher H., Mundlos S., Winterpacht A., Zabel B.,
RA Fidler A., Gruss P., Plachov D.;
RT "PAX8, a human paired box gene; isolation and expression in developing thyroid, kidney and Wilms' tumors.";
RL Development 116:611-623(1992).
```

[2] SEQUENCE FROM N.A. (ISOFORM 1), AND ALTERNATIVE SPLICING.
RP TISSUE=Kidney;
RX MEDLINE=94015277; PubMed=8413205;
RA Kozmik Z., Kurzbauser R., Doerfler P., Busslinger M.;
RT "Alternative splicing of Pax-8 gene transcripts is developmentally
regulated and generates isoforms with different transactivation
properties.";
RL Mol. Cell. Biol. 13:6024-6035(1993).
[3]
[4] SEQUENCE FROM N.A. (ISOFORMS 1; 3; 4 AND 5).
RX MEDLINE=95255300; PubMed=773192;
RA Poleev A., Wendler F., Fickenscher H., Zamini M.S., Yeginuma K.,
Abbott C., Plachov D.;
RT "Distinct functional properties of three human paired-box-protein,
PAX8, isoforms generated by alternative splicing in thyroid, kidney
and Wilms' tumors.";
RL Eur. J. Biochem. 228:899-911(1995).
[5] VARIANT TD PRO-40.
RX MEDLINE=21394123; PubMed=11502839;
RA Congdon T., Nguyen L.Q., Nogueira C.R., Habiby R.L., Medeiros-Neto G.,
Kopp P.;
RT "A novel mutation (Q40P) in PAX8 associated with congenital
hypothyroidism and thyroid hypoplasia: evidence for phenotypic
variability in mother and child.";
RL J. Clin. Endocrinol. Metab. 86:3962-3967(2001).
[6] VARIANTS TD HIS-31 AND ARG-62.
RX MEDLINE=98250178; PubMed=9590296;
RA Macchia P.E., Lapi P., Krude H., Pizzo M.T., Missero C., Chiavato L.,
Souabni A., Baserga M., Tassi V., Pinchera A., Penzi G., Gruters A.,
Busslinger M., Di Lauro R.;
RT "PAX8 mutations associated with congenital hypothyroidism caused by
thyroid dysgenesis.";
RL Nat. Genet. 19:83-86(1998).
[7] VARIANT TD TYR-57.
RX MEDLINE=21142581; PubMed=11232006;
RA Vallin C., Rydlewski C., Duprez L., Heinrichs C., Abramowicz M.,
Malvaux P., Renneboog B., Parma J., Costagliola S., Vassart G.;
RT "Autosomal dominant transmission of congenital thyroid hypoplasia due
to loss-of-function mutation of PAX8.";
RL J. Clin. Endocrinol. Metab. 86:234-238(2001).
[8] -1- FUNCTION: Transcription factor for the thyroid-specific expression
of the genes exclusively expressed in the thyroid cell type,
maintaining the functional differentiation of such cells.
[9] -1- SUBCELLULAR LOCATION: Nuclear.
[10] -1- ALTERNATIVE PRODUCTS:
Name=1; Synonyms=Pax8a;
IsoId=Q06710-1; Sequence=Displayed;
Name=2; Synonyms=Pax8b;
IsoId=Q06710-2; Sequence=VSP_002372;
Name=3; Synonyms=Pax8c;
IsoId=Q06710-3; Sequence=VSP_002373;
Name=4; Synonyms=Pax8d;
IsoId=Q06710-4; Sequence=VSP_002374;
Name=5; Synonyms=Pax8e;
IsoId=Q06710-5; Sequence=VSP_002375;
[11] TISSUE SPECIFICITY: EXPRESSED IN THE EXCRETORY SYSTEM, THYROID
GLAND AND WILMS TUMORS.
[12] -1- DEVELOPMENTAL STAGE: IN DEVELOPING EXCRETORY SYSTEM, DURING
THYROID DIFFERENTIATION AND IN ADULT THYROID.
[13] -1- DISEASE: Defects in PAX8 are a cause of thyroid dysgenesis (TD),
[MIM:218700]; the most frequent cause of congenital hypothyroidism
(CH). The thyroid gland can be absent (athyreosis), ectopically
located (most often at the base of the tongue) and/or severely
reduced in size (hypoplasia).
[14] -1- SIMILARITY: Contains 1 paired box domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC NOT ANNOTATED_CDS.
DR EMBL; X69699; -; -;
DR EMBL; L19606; AAA03539.1; -;
DR EMBL; S77904; AAB34216.1; -;
DR EMBL; S77905; AAB34217.2; -;
DR EMBL; S77906; AAB34218.2; -;
DR HSSP; P26367; 6PAX.
DR TRANSFAC; T01824; -;
DR TRANSFAC; T01825; -;
DR TRANSFAC; T01826; -;
DR TRANSFAC; T01827; -;
DR TRANSFAC; T02898; -;
DR Gensw; HGNC:8622; PAX8.
DR MIM; 167415; -;
DR MIM; 218700; -;
DR GO; GO:0005634; C:nucleus; NAS.
DR GO; GO:0004996; F:thyroid-stimulating hormone receptor activity; TAS.
DR GO; GO:0003700; F:transcription factor activity; TAS.
DR GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.
DR GO; GO:0007397; P:histogenesis and organogenesis; TAS.
DR InterPro; IPR001523; Paired_box.
DR Pfam; PF00292; PAX; 1.
DR PRINTS; PRO0027; PAIREDBOX.
DR SMART; SM00351; PAX; 1.
DR PROSITE; PS00334; PAIRED_BOX; 1.
KW Paired box; Developmental protein; Nuclear protein; DNA-binding;
KW Transcription regulation; Differentiation; Alternative splicing;
KW Disease mutation; Polymorphism.
FT DOMAIN 9 133 PAIRED_BOX.
FT DOMAIN 315 327
FT VARSPPLIC 300 362
FT VARSPPLIC 301 450
FT VARSPPLIC 261 450
FT VARSPPLIC 260 450
FT VARSPPLIC 31 31
FT VARSPPLIC 40 40
FT VARSPPLIC 57 57
FT VARSPPLIC 62 62
FT VARSPPLIC 329 329
FT VARSPPLIC 300 300
FT VARSPPLIC 418 418
FT VARSPPLIC 450 AA; 4811 MW; 626C21FCEAF20A0E CRC64;
SEQUENCE 450 AA; 4811 MW; 626C21FCEAF20A0E CRC64;

CC the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC NOT ANNOTATED_CDS.
DR EMBL; X69699; -; -;
DR EMBL; L19606; AAA03539.1; -;
DR EMBL; S77904; AAB34216.1; -;
DR EMBL; S77905; AAB34217.2; -;
DR EMBL; S77906; AAB34218.2; -;
DR HSSP; P26367; 6PAX.
DR TRANSFAC; T01824; -;
DR TRANSFAC; T01825; -;
DR TRANSFAC; T01826; -;
DR TRANSFAC; T01827; -;
DR TRANSFAC; T02898; -;
DR Gensw; HGNC:8622; PAX8.
DR MIM; 167415; -;
DR MIM; 218700; -;
DR GO; GO:0005634; C:nucleus; NAS.
DR GO; GO:0004996; F:thyroid-stimulating hormone receptor activity; TAS.
DR GO; GO:0003700; F:transcription factor activity; TAS.
DR GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.
DR GO; GO:0007397; P:histogenesis and organogenesis; TAS.
DR InterPro; IPR001523; Paired_box.
DR Pfam; PF00292; PAX; 1.
DR PRINTS; PRO0027; PAIREDBOX.
DR SMART; SM00351; PAX; 1.
DR PROSITE; PS00334; PAIRED_BOX; 1.
KW Paired box; Developmental protein; Nuclear protein; DNA-binding;
KW Transcription regulation; Differentiation; Alternative splicing;
KW Disease mutation; Polymorphism.
FT DOMAIN 9 133 PAIRED_BOX.
FT DOMAIN 315 327
FT VARSPPLIC 300 362
FT VARSPPLIC 301 450
FT VARSPPLIC 261 450
FT VARSPPLIC 260 450
FT VARSPPLIC 31 31
FT VARSPPLIC 40 40
FT VARSPPLIC 57 57
FT VARSPPLIC 62 62
FT VARSPPLIC 329 329
FT VARSPPLIC 300 300
FT VARSPPLIC 418 418
FT VARSPPLIC 450 AA; 4811 MW; 626C21FCEAF20A0E CRC64;
SEQUENCE 450 AA; 4811 MW; 626C21FCEAF20A0E CRC64;

CC the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC NOT ANNOTATED_CDS.
DR EMBL; X69699; -; -;
DR EMBL; L19606; AAA03539.1; -;
DR EMBL; S77904; AAB34216.1; -;
DR EMBL; S77905; AAB34217.2; -;
DR EMBL; S77906; AAB34218.2; -;
DR HSSP; P26367; 6PAX.
DR TRANSFAC; T01824; -;
DR TRANSFAC; T01825; -;
DR TRANSFAC; T01826; -;
DR TRANSFAC; T01827; -;
DR TRANSFAC; T02898; -;
DR Gensw; HGNC:8622; PAX8.
DR MIM; 167415; -;
DR MIM; 218700; -;
DR GO; GO:0005634; C:nucleus; NAS.
DR GO; GO:0004996; F:thyroid-stimulating hormone receptor activity; TAS.
DR GO; GO:0003700; F:transcription factor activity; TAS.
DR GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.
DR GO; GO:0007397; P:histogenesis and organogenesis; TAS.
DR InterPro; IPR001523; Paired_box.
DR Pfam; PF00292; PAX; 1.
DR PRINTS; PRO0027; PAIREDBOX.
DR SMART; SM00351; PAX; 1.
DR PROSITE; PS00334; PAIRED_BOX; 1.
KW Paired box; Developmental protein; Nuclear protein; DNA-binding;
KW Transcription regulation; Differentiation; Alternative splicing;
KW Disease mutation; Polymorphism.
FT DOMAIN 9 133 PAIRED_BOX.
FT DOMAIN 315 327
FT VARSPPLIC 300 362
FT VARSPPLIC 301 450
FT VARSPPLIC 261 450
FT VARSPPLIC 260 450
FT VARSPPLIC 31 31
FT VARSPPLIC 40 40
FT VARSPPLIC 57 57
FT VARSPPLIC 62 62
FT VARSPPLIC 329 329
FT VARSPPLIC 300 300
FT VARSPPLIC 418 418
FT VARSPPLIC 450 AA; 4811 MW; 626C21FCEAF20A0E CRC64;
SEQUENCE 450 AA; 4811 MW; 626C21FCEAF20A0E CRC64;

CC the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC NOT ANNOTATED_CDS.
DR EMBL; X69699; -; -;
DR EMBL; L19606; AAA03539.1; -;
DR EMBL; S77904; AAB34216.1; -;
DR EMBL; S77905; AAB34217.2; -;
DR EMBL; S77906; AAB34218.2; -;
DR HSSP; P26367; 6PAX.
DR TRANSFAC; T01824; -;
DR TRANSFAC; T01825; -;
DR TRANSFAC; T01826; -;
DR TRANSFAC; T01827; -;
DR TRANSFAC; T02898; -;
DR Gensw; HGNC:8622; PAX8.
DR MIM; 167415; -;
DR MIM; 218700; -;
DR GO; GO:0005634; C:nucleus; NAS.
DR GO; GO:0004996; F:thyroid-stimulating hormone receptor activity; TAS.
DR GO; GO:0003700; F:transcription factor activity; TAS.
DR GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.
DR GO; GO:0007397; P:histogenesis and organogenesis; TAS.
DR InterPro; IPR001523; Paired_box.
DR Pfam; PF00292; PAX; 1.
DR PRINTS; PRO0027; PAIREDBOX.
DR SMART; SM00351; PAX; 1.
DR PROSITE; PS00334; PAIRED_BOX; 1.
KW Paired box; Developmental protein; Nuclear protein; DNA-binding;
KW Transcription regulation; Differentiation; Alternative splicing;
KW Disease mutation; Polymorphism.
FT DOMAIN 9 133 PAIRED_BOX.
FT DOMAIN 315 327
FT VARSPPLIC 300 362
FT VARSPPLIC 301 450
FT VARSPPLIC 261 450
FT VARSPPLIC 260 450
FT VARSPPLIC 31 31
FT VARSPPLIC 40 40
FT VARSPPLIC 57 57
FT VARSPPLIC 62 62
FT VARSPPLIC 329 329
FT VARSPPLIC 300 300
FT VARSPPLIC 418 418
FT VARSPPLIC 450 AA; 4811 MW; 626C21FCEAF20A0E CRC64;
SEQUENCE 450 AA; 4811 MW; 626C21FCEAF20A0E CRC64;

CC the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC NOT ANNOTATED_CDS.
DR EMBL; X69699; -; -;
DR EMBL; L19606; AAA03539.1; -;
DR EMBL; S77904; AAB34216.1; -;
DR EMBL; S77905; AAB34217.2; -;
DR EMBL; S77906; AAB34218.2; -;
DR HSSP; P26367; 6PAX.
DR TRANSFAC; T01824; -;
DR TRANSFAC; T01825; -;
DR TRANSFAC; T01826; -;
DR TRANSFAC; T01827; -;
DR TRANSFAC; T02898; -;
DR Gensw; HGNC:8622; PAX8.
DR MIM; 167415; -;
DR MIM; 218700; -;
DR GO; GO:0005634; C:nucleus; NAS.
DR GO; GO:0004996; F:thyroid-stimulating hormone receptor activity; TAS.
DR GO; GO:0003700; F:transcription factor activity; TAS.
DR GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.
DR GO; GO:0007397; P:histogenesis and organogenesis; TAS.
DR InterPro; IPR001523; Paired_box.
DR Pfam; PF00292; PAX; 1.
DR PRINTS; PRO0027; PAIREDBOX.
DR SMART; SM00351; PAX; 1.
DR PROSITE; PS00334; PAIRED_BOX; 1.
KW Paired box; Developmental protein; Nuclear protein; DNA-binding;
KW Transcription regulation; Differentiation; Alternative splicing;
KW Disease mutation; Polymorphism.
FT DOMAIN 9 133 PAIRED_BOX.
FT DOMAIN 315 327
FT VARSPPLIC 300 362
FT VARSPPLIC 301 450
FT VARSPPLIC 261 450
FT VARSPPLIC 260 450
FT VARSPPLIC 31 31
FT VARSPPLIC 40 40
FT VARSPPLIC 57 57
FT VARSPPLIC 62 62
FT VARSPPLIC 329 329
FT VARSPPLIC 300 300
FT VARSPPLIC 418 418
FT VARSPPLIC 450 AA; 4811 MW; 626C21FCEAF20A0E CRC64;
SEQUENCE 450 AA; 4811 MW; 626C21FCEAF20A0E CRC64;

CC the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC NOT ANNOTATED_CDS.
DR EMBL; X69699; -; -;
DR EMBL; L19606; AAA03539.1; -;
DR EMBL; S77904; AAB34216.1; -;
DR EMBL; S77905; AAB34217.2; -;
DR EMBL; S77906; AAB34218.2; -;
DR HSSP; P26367; 6PAX.
DR TRANSFAC; T01824; -;
DR TRANSFAC; T01825; -;
DR TRANSFAC; T01826; -;
DR TRANSFAC; T01827; -;
DR TRANSFAC; T02898; -;
DR Gensw; HGNC:8622; PAX8.
DR MIM; 167415; -;
DR MIM; 218700; -;
DR GO; GO:0005634; C:nucleus; NAS.
DR GO; GO:0004996; F:thyroid-stimulating hormone receptor activity; TAS.
DR GO; GO:0003700; F:transcription factor activity; TAS.
DR GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.
DR GO; GO:0007397; P:histogenesis and organogenesis; TAS.
DR InterPro; IPR001523; Paired_box.
DR Pfam; PF00292; PAX; 1.
DR PRINTS; PRO0027; PAIREDBOX.
DR SMART; SM00351; PAX; 1.
DR PROSITE; PS00334; PAIRED_BOX; 1.
KW Paired box; Developmental protein; Nuclear protein; DNA-binding;
KW Transcription regulation; Differentiation; Alternative splicing;
KW Disease mutation; Polymorphism.
FT DOMAIN 9 133 PAIRED_BOX.
FT DOMAIN 315 327
FT VARSPPLIC 300 362
FT VARSPPLIC 301 450
FT VARSPPLIC 261 450
FT VARSPPLIC 260 450
FT VARSPPLIC 31 31
FT VARSPPLIC 40 40
FT VARSPPLIC 57 57
FT VARSPPLIC 62 62
FT VARSPPLIC 329 329
FT VARSPPLIC 300 300
FT VARSPPLIC 418 418
FT VARSPPLIC 450 AA; 4811 MW; 626C21FCEAF20A0E CRC64;
SEQUENCE 450 AA; 4811 MW; 626C21FCEAF20A0E CRC64;

CC the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC NOT ANNOTATED_CDS.
DR EMBL; X69699; -; -;
DR EMBL; L19606; AAA03539.1; -;
DR EMBL; S77904; AAB34216.1; -;
DR EMBL; S77905; AAB34217.2; -;
DR EMBL; S77906; AAB34218.2; -;
DR HSSP; P26367; 6PAX.
DR TRANSFAC; T01824; -;
DR TRANSFAC; T01825; -;
DR TRANSFAC; T01826; -;
DR TRANSFAC; T01827; -;
DR TRANSFAC; T02898; -;
DR Gensw; HGNC:8622; PAX8.
DR MIM; 167415; -;
DR MIM; 218700; -;
DR GO; GO:0005634; C:nucleus; NAS.
DR GO; GO:0004996; F:thyroid-stimulating hormone receptor activity; TAS.
DR GO; GO:0003700; F:transcription factor activity; TAS.
DR GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.
DR GO; GO:0007397; P:histogenesis and organogenesis; TAS.
DR InterPro; IPR001523; Paired_box.
DR Pfam; PF00292; PAX; 1.
DR PRINTS; PRO0027; PAIREDBOX.
DR SMART; SM00351; PAX; 1.
DR PROSITE; PS00334; PAIRED_BOX; 1.
KW Paired box; Developmental protein; Nuclear protein; DNA-binding;
KW Transcription regulation; Differentiation; Alternative splicing;
KW Disease mutation; Polymorphism.
FT DOMAIN 9 133 PAIRED_BOX.
FT DOMAIN 315 327
FT VARSPPLIC 300 362
FT VARSPPLIC 301 450
FT VARSPPLIC 261 450
FT VARSPPLIC 260 450
FT VARSPPLIC 31 31
FT VARSPPLIC 40 40
FT VARSPPLIC 57 57
FT VARSPPLIC 62 62
FT VARSPPLIC 329 329
FT VARSPPLIC 300 300
FT VARSPPLIC 418 418
FT VARSPPLIC 450 AA; 4811 MW; 626C21FCEAF20A0E CRC64;
SEQUENCE 450 AA; 4811 MW; 626C21FCEAF20A0E CRC64;

CC the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC NOT ANNOTATED_CDS.
DR EMBL; X69699; -; -;
DR EMBL; L19606; AAA03539.1; -;
DR EMBL; S77904; AAB34216.1; -;
DR EMBL; S77905; AAB34217.2; -;
DR EMBL; S77906; AAB34218.2; -;
DR HSSP; P26367; 6PAX.
DR TRANSFAC; T01824; -;
DR TRANSFAC; T01825; -;
DR TRANSFAC; T01826; -;
DR TRANSFAC; T01827; -;
DR TRANSFAC; T02898; -;
DR Gensw; HGNC:8622; PAX8.
DR MIM; 167415; -;
DR MIM; 218700; -;
DR GO; GO:0005634; C:nucleus; NAS.
DR GO; GO:0004996; F:thyroid-stimulating hormone receptor activity; TAS.
DR GO; GO:0003700; F:transcription factor activity; TAS.
DR GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.
DR GO; GO:0007397; P:histogenesis and organogenesis; TAS.
DR InterPro; IPR001523; Paired_box.
DR Pfam; PF00292; PAX; 1.
DR PRINTS; PRO0027; PAIREDBOX.
DR SMART; SM00351; PAX; 1.
DR PROSITE; PS00334; PAIRED_BOX; 1.
KW Paired box; Developmental protein; Nuclear protein; DNA-binding;
KW Transcription regulation; Differentiation; Alternative splicing;
KW Disease mutation; Polymorphism.
FT DOMAIN 9 133 PAIRED_BOX.
FT DOMAIN 315 327
FT VARSPPLIC 300 362
FT VARSPPLIC 301 450
FT VARSPPLIC 261 450
FT VARSPPLIC 260 450
FT VARSPPLIC 31 31
FT VARSPPLIC 40 40
FT VARSPPLIC 57 57
FT VARSPPLIC 62 62
FT VARSPPLIC 329 329
FT VARSPPLIC 300 300
FT VARSPPLIC 418 418
FT VARSPPLIC 450 AA; 4811 MW; 626C21FCEAF20A0E CRC64;
SEQUENCE 450 AA; 4811 MW; 626C21FCEAF20A0E CRC64;

CC the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC NOT ANNOTATED_CDS.
DR EMBL; X69699; -; -;
DR EMBL; L19606; AAA03539.1; -;
DR EMBL; S77904; AAB34216.1; -;
DR EMBL; S77905; AAB34217.2; -;
DR EMBL; S77906; AAB34218.2; -;
DR HSSP; P26367; 6PAX.
DR TRANSFAC; T01824; -;
DR TRANSFAC; T01825; -;
DR TRANSFAC; T01826; -;
DR TRANSFAC; T01827; -;
DR TRANSFAC; T02898; -;
DR Gensw; HGNC:8622; PAX8.
DR MIM; 167415; -;
DR MIM; 218700; -;
DR GO; GO:0005634; C:nucleus; NAS.
DR GO; GO:0004996; F:thyroid-stimulating hormone receptor activity; TAS.
DR GO; GO:0003700; F:transcription factor activity; TAS.
DR GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.
DR GO; GO:0007397; P:histogenesis and organogenesis; TAS.
DR InterPro; IPR001523; Paired_box.
DR Pfam; PF00292; PAX; 1.
DR PRINTS; PRO0027; PAIREDBOX.
DR SMART; SM00351; PAX; 1.
DR PROSITE; PS00334; PAIRED_BOX; 1.
KW Paired box; Developmental protein; Nuclear protein; DNA-binding;
KW Transcription regulation; Differentiation; Alternative splicing;
KW Disease mutation; Polymorphism.
FT DOMAIN 9 133 PAIRED_BOX.
FT DOMAIN 315 327
FT VARSPPLIC 300 362
FT VARSPPLIC 301 450
FT VARSPPLIC 261 450
FT VARSPPLIC 260 450
FT VARSPPLIC 31 31
FT VARSPPLIC 40 40
FT VARSPPLIC 57 57
FT VARSPPLIC 62 62
FT VARSPPLIC 329 329
FT VARSPPLIC 300 300
FT VARSPPLIC 418 418
FT VARSPPLIC 450 AA; 4811 MW; 626C21FCEAF20A0E CRC64;
SEQUENCE 450 AA; 4811 MW; 626C21FCEAF20A0E CRC64;

CC the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC NOT ANNOTATED_CDS.
DR EMBL; X69699; -; -;
DR EMBL; L19606; AAA03539.1; -;
DR EMBL; S77904; AAB34216.1; -;
DR EMBL; S77905; AAB34217.2; -;
DR EMBL; S77906; AAB34218.2; -;
DR HSSP; P26367; 6PAX.
DR TRANSFAC; T01824; -;
DR TRANSFAC; T01825; -;
DR TRANSFAC; T01826; -;
DR TRANSFAC; T01827; -;
DR TRANSFAC; T02898; -;
DR Gensw; HGNC:8622; PAX8.
DR MIM; 167415; -;
DR MIM; 218700; -;
DR GO; GO:0005634; C:nucleus; NAS.
DR GO; GO:0004996; F:thyroid-stimulating hormone receptor activity; TAS.
DR GO; GO:0003700; F:transcription factor activity; TAS.
DR GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.
DR GO; GO:0007397; P:histogenesis and organogenesis; TAS.
DR InterPro; IPR001523; Paired_box.
DR Pfam; PF00292; PAX; 1.
DR PRINTS; PRO0027; PAIREDBOX.
DR SMART; SM00351; PAX; 1.
DR PROSITE; PS00334; PAIRED_BOX; 1.
KW Paired box; Developmental protein; Nuclear protein; DNA-binding;
KW Transcription regulation; Differentiation; Alternative splicing;
KW Disease mutation; Polymorphism.
FT DOMAIN 9 133 PAIRED_BOX.
FT DOMAIN 315 327
FT VARSPPLIC 300 362
FT VARSPPLIC 301 450
FT VARSPPLIC 261 450
FT VARSPPLIC 260 450
FT VARSPPLIC 31 31
FT VARSPPLIC 40 40
FT VARSPPLIC 57 57
FT VARSPPLIC 62 62
FT VARSPPLIC 329 329
FT VARSPPLIC 300 300
FT VARSPPLIC 418 418
FT VARSPPLIC 450 AA; 4811 MW; 626C21FCEAF20A0E CRC64;
SEQUENCE 450 AA; 4811 MW; 626C21FCEAF20A0E CRC64;

CC the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC NOT ANNOTATED_CDS.
DR EMBL; X69699; -; -;
DR EMBL; L19606; AAA03539.1; -;
DR EMBL; S77904; AAB34216.1; -;
DR EMBL; S77905; AAB34217.2; -;
DR EMBL; S77906; AAB34218.2; -;
DR HSSP; P26367; 6PAX.
DR TRANSFAC; T01824; -;
DR TRANSFAC; T01825; -;
DR TRANSFAC; T01826; -;
DR TRANSFAC; T01827; -;
DR TRANSFAC; T02898; -;
DR Gensw; HGNC:8622; PAX8.
DR MIM; 167415; -;
DR MIM; 218700; -;
DR GO; GO:0005634; C:nucleus; NAS.
DR GO; GO:0004996; F:thyroid-stimulating hormone receptor activity; TAS.
DR GO; GO:0003700; F:transcription factor activity; TAS.
DR GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.
DR GO; GO:0007397; P:histogenesis and organogenesis; TAS.
DR InterPro; IPR001523; Paired_box.
DR Pfam; PF00292; PAX; 1.
DR PRINTS; PRO0027; PAIREDBOX.
DR SMART; SM00351; PAX; 1.
DR PROSITE; PS00334; PAIRED_BOX; 1.
KW Paired box; Developmental protein; Nuclear protein; DNA-binding;
KW Transcription regulation; Differentiation; Alternative splicing;
KW Disease mutation; Polymorphism.
FT DOMAIN 9 133 PAIRED_BOX.
FT DOMAIN 315 327
FT VARSPPLIC 300 362
FT VARSPPLIC 301 450
FT VARSPPLIC 261 450
FT VARSPPLIC 260 450
FT VARSPPLIC 31 31
FT VARSPPLIC 40 40
FT VARSPPLIC 57 57
FT VARSPPLIC 62 62
FT VARSPPLIC 329 329
FT VARSPPLIC 300 300
FT VARSPPLIC 418 418
FT VARSPPLIC 450 AA; 4811 MW; 626C21FCEAF20A0E CRC64;
SEQUENCE 450 AA; 4811 MW; 626C21FCEAF20A0E CRC64;

CC the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC NOT ANNOTATED_CDS.
DR EMBL; X69699; -; -;
DR EMBL; L19606; AAA03539.1; -;
DR EMBL; S77904; AAB34216.1; -;
DR EMBL; S77905; AAB34217.2; -;
DR EMBL; S77906; AAB34218.2; -;
DR HSSP; P26367; 6PAX.
DR TRANSFAC; T01824; -;
DR TRANSFAC; T01825; -;
DR TRANSFAC; T01826; -;
DR TRANSFAC; T01827; -;
DR TRANSFAC; T02898; -;
DR Gensw; HGNC:8622; PAX8.
DR MIM; 167415; -;
DR MIM; 218700; -;
DR GO; GO:0005634; C:nucleus; NAS.
DR GO; GO:0004996; F:thyroid-stimulating hormone receptor activity; TAS.
DR GO; GO:0003700; F:transcription factor activity; TAS.
DR GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.
DR GO; GO:0007397; P:histogenesis and organogenesis; TAS.
DR InterPro; IPR001523; Paired_box.
DR Pfam; PF00292; PAX; 1.
DR PRINTS; PRO0027; PAIREDBOX.
DR SMART; SM00351; PAX; 1.
DR PROSITE; PS00334; PAIRED_BOX; 1.
KW Paired box; Developmental protein; Nuclear protein; DNA-binding;
KW Transcription regulation; Differentiation; Alternative splicing;
KW Disease mutation; Polymorphism.
FT DOMAIN 9 133 PAIRED_BOX.
FT DOMAIN 315 327
FT VARSPPLIC 300 362
FT VARSPPLIC 301 450
FT VARSPPLIC 261 450
FT VARSPPLIC 260 450
FT VARSPPLIC 31 31
FT VARSPPLIC 40 40
FT VARSPPLIC 57 57
FT VARSPPLIC 62 62
FT VARSPPLIC 329 329
FT VARSPPLIC 300 300
FT VARSPPLIC 418 418
FT VARSPPLIC 450 AA; 4811 MW; 626C21FCEAF20A0E CRC64;
SEQU

Query Match 4.9%; Score 105; DB 1; Length 450;
Best Local Similarity 18.6%; Pred. No. 0.62;
Matches 80; Conservative 51; Mismatches 125; Indels 174; Gaps 17;
QY 4 PELKESVIVSRHGVRAPTKATQL-----MQDVTDPAMVTPVKLGEL-----TP 49
DB 27 PEWQRIVDLAQGVPCDISRLRVSHGCVSKILGRYETGRIRPGVIGSKPKVATP 86
QY 50 RGELIAYLGHWRQ-----RLVADGL-----LP 73
DB 87 K--VVEKIDYKRONPTWFAEIRLLAEGVCDNTVPVSSINRIIRTKVQPPNLP 143
QY 74 KCGC-----POSGQVAIIADVDERTRKTGEAFA-----AGLA----- 105
DB 144 MDCSVATKSLSPCHTILPSS---AVTPSPQSDSLGTSYINGLLGIAQPGSKDKKMD 200
QY 106 ---PDAITVHQADTSSPDPLFNPLKTVGVCOLDNANVTDAILERAGGSIADFTGHYOTA 162
DB 201 SDQDSCLSIDSQSSSGPR---KHLRTDAFQSHLEPLECPFERQ-----HYPEA 248
QY 163 PRELERVLPNPSQCLKREKQDECSLTQALPSELKVSADCVSL-----TGAVSILA 214
DB 249 Y-----APSHTKGEQGLYPLPLNLSLTDGKATLT 279
QY 215 SMLTEIFLLQQAQGMPEPCWGRITDSHOWNTLLSHNAQFDLLQRTPEVARSATP----- 270
DB 280 PSNTPL-----GRNLSHTQYVPVADPHSPFAIKQETPEVSSSSSTPSSLS 325
QY 271 ---LDDLIKALTTPHPQQAQGVTLPTSVLF-----IAGHDTNLANLGGALELAWTLPG 322
DB 326 SSAFLDL-----QQVSGVPPFNAPPHAASVYQGTGALLSGREKVGFTLPG 373
QY 323 QPDNTPPGGE 332
DB 374 YPHIPTSGQ 383
RESULT 13
PPAL MOUSE
ID PPAL MOUSE STANDARD; PRT; 423 AA.
AC P24638; Q8QZT5;
DT 01-MAR-1992 (Rel. 21, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Lysosomal acid phosphatase precursor (EC 3.1.3.2) (LAP).
GN ACP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RP TISSUE=Liver;
RC MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.P., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Guraratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahney J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalish D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
"Generation and initial analysis of more than 15,000 full-length human

and mouse cDNA sequences.";
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[2]
SEQUENCE OF 3-423 FROM N.A.
MEDLINE=91282986; PubMed=2059337;
RA Geier C., von Figura K., Pohlmann R.;
RT "Molecular cloning of the mouse lysosomal acid phosphatase.";
Biol. Chem. Hoppe-Seyler 372:301-304(1991).
CC -!- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O = an
CC alcohol + phosphate.
CC -!- SUBCELLULAR LOCATION: Lysosomal.
CC -!- SIMILARITY: Belongs to the histidine acid phosphatase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; BC023343; AAH23343.1; -.
DR EMBL; X57199; CAA40485.1; -.
DR PIR; S14742; S14742.
DR HSSP; P15309; ZHPA..
DR MGD; MGI:87882; ACP2.
DR InterPro; IPR000560; HisAc phsphtse.
DR Pfam; PF00328; acid phosphat; 1.
DR PROSITE; PS00616; HIS ACID PHOSPHAT 1; 1.
DR PROSITE; PS00778; HIS ACID PHOSPHAT 2; 1.
DR Hydrolase; Signal; Glycoprotein; Lysosome.
KW SIGNAL 1 30
FT CHAIN 31 423 LYSOSOMAL ACID PHOSPHATASE.
FT DISULFID 159 370 BY SIMILARITY.
FT DISULFID 212 310 BY SIMILARITY.
FT DISULFID 345 349 BY SIMILARITY.
FT ACT SITE 42 42 NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).
FT ACT SITE 84 84 BY SIMILARITY.
FT ACT SITE 286 286 PROTON DONOR (BY SIMILARITY).
FT CARBOHYD 92 92 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 167 167 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 177 177 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 197 197 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 267 267 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 374 374 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 131 131 N -> S (IN REF. 2).
SQ SEQUENCE 423 AA; 48508 MW; AFD7A5C90A4F2AF4 CRC64;
Query Match 4.8%; Score 104; DB 1; Length 423;
Best Local Similarity 21.4%; Pred. No. 0.69; Mismatches 122; Gaps 21;
Matches 91; Conservative 47;
QY 8 LESVIVSRHGVRAPTKATQIMQDVTDPAMVTPVKLGELTPRGELIAYLGHYWRQLV 67
DB 33 LRFVTLVYRHGDRSPVKT----YPKDFYQBEKMPQGFQLTKEGMLQHWELGQALRQRY- 87
QY 68 ADGLLPKCGCQSQQVAIIADV-ERTRKGEAFAAGLAPDCAITVQTADTSSPDPLFN 126
DB 88 -HGFL---NTSYHRQEVYVRSTDFDRTLMSEANLGLFPP-----NEVOHFNPNISWQ 137
QY 127 PLKTVGVCOLDNANVTDAILERAGGSIADFTGHYQTAFAERLERVLNFFQSNLCLKREKQDE 186
DB 138 PI-----PVHTVPITE-----DRLKFLPGPCPRYEQLQNE 168
QY 187 SCSTQALPSELKVSADCVSLTGVASLMTETFL-----LQQAQGMPEPGWG 235
DB 169 ---TRQTPPEYQNRISQNAQFLNANVANETGLNTVLETLNWNVYDTLFCQTHGLLPPWA 224
QY 236 -----RITDSSOWN--TLLSLHNAQFDLLQRTPEVARSR-----ATPLDLIKALT 280

```

Db      225 SPQTQVRISQUNDKFSLFLFGH-----QVKARLQGVLLAQILKNLTMA 274
Qy      281 PHPPQKQAYGVLTPTSVLFIAGHTDNLNLANLGALNWTLPQPDNTPPGGELVFERMR 340
Db      275 SQPKK-----LLVYSAHDTTVALQWALNVY-----NGKQAPYASCHIFELYQ- 317
Qy      341 LSNDSQWTSQVSLVFTQLQOMKDKTFLSLNTPPGVEKVLTLAGCEERNAQGWCSLAGTQIV 400
Db      318 -EDGNF-SVENYFRNDSK-----KAPWPLILP-----GCPHR-----CPLQDFLRL- 357
Qy      401 NEARIP 406
Db      358 TEPIVP 363

RESULT 14
ID DIVL CAUCR
ID DIVL CAUCR STANDARD; PRT; 769 AA.
AC Q9RQ09; Q9A2S2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Sensor protein divL (EC 2.7.3.-).
GN DIVL OR CC3484.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
OC Caulobacteraceae; Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
SEQUENCE FROM N.A., AND AUTOPHOSPHORYLATION SITE.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=20027501; PubMed=10557274;
RA Wu J., Ohta N., Zhao J.B., Newton A.;
RT "A novel bacterial tyrosine kinase essential for cell division and
RL differentiation.";
RR Proc. Natl. Acad. Sci. U.S.A. 96:13068-13073(1999).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nieman W.C. Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.B.,
RA Eisen J.A., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R., Ely B.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.B., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uterback T., Tran K., Wolf A., Vamathavan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RC Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
CC -1- FUNCTION: Required for cell division and growth. It catalyzes the
CC phosphorylation of ttrA and activates transcription in vitro of
CC the cell cycle-regulated flif promoter.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- PTM: Autophosphorylated.
CC -1- SIMILARITY: Contains 1 histidine kinase domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announcement
CC or send an email to license@isb-sib.ch).
CC -----
CC ENBL; AF083422; AAP08344.2; -
CC DR ENBL; AF006007; ABK25446.1; -
CC DR TIGR; B87681; B87681.
CC DR TIGR; CC3484; -.
CC DR PhosSite; Q9RQ09; -.
CC DR InterPro; IPR003594; ATPbind_Atpase.
CC DR InterPro; IPR004358; Bact_sens_pr_C.
CC DR InterPro; IPR003661; His_Kin_N_.
CC -----

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 30, 2004, 12:34:35 ; Search time 37.1429 Seconds
(without alignments)
3482.836 Million cell updates/sec

Title: US-10-021-723B-13
Perfect score: 2153
Sequence: 1 QSEPELKESVIVSRHGVR.....CSLAGFTQIVNEARIPACSL 410

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL 25:*

- 1: sp_archea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2073	96.3	432	2 Q8GN88	Q8gn88 escherichia
2	2071	96.2	432	2 Q8KDK6	Q8rkd6 escherichia
3	2071	96.2	432	2 Q8KDK7	Q8rkd7 escherichia
4	2071	96.2	432	2 Q8KDK8	Q8rkd8 escherichia
5	2071	96.2	432	2 Q8KDK9	Q8rkd9 escherichia
6	2070	96.1	432	2 Q8KDK5	Q8rkd5 escherichia
7	2067	96.0	432	2 Q8KDK3	Q8rkd3 shigella fl
8	2044	94.9	442	16 Q83RW2	Q83rw2 shigella fl
9	2042	94.8	434	16 Q8XC29	Q8xc29 escherichia
10	2042	94.8	446	16 Q8CW75	Q8cw75 escherichia
11	2040	94.8	432	16 Q7UD08	Q7ud08 shigella fl
12	954	44.3	441	16 Q8ZFP6	Q8zfp6 versinia de
13	551	25.6	413	16 Q8AB26	Q8ab26 escherichia
14	543	25.2	413	16 Q8Z7F1	Q8z7fp salmonella
15	543	25.2	413	16 Q7UD02	Q7ud02 shigella fl
16	543	25.2	421	16 Q83RV6	Q83rv6 shigella fl

17	534.5	24.8	414	16	Q9AAQ4	Q9aaq4 caulobacter
18	511	23.7	392	16	Q8PP76	Q8pp76 xanthomonas
19	497	23.1	443	16	Q8P330	Q8p330 xanthomonas
20	455.5	21.2	435	16	Q8PF53	Q8pf53 xanthomonas
21	432	20.1	428	2	Q8GD20	Q8gd20 pseudomonas
22	394	18.3	318	16	Q8CW72	Q8cw72 escherichia
23	357.5	16.6	414	16	Q8PPX6	Q8ppx6 xanthomonas
24	356	16.5	421	2	Q84CN9	Q84cn9 klebsiella
25	348	16.2	421	2	Q7WSY1	Q7wsy1 klebsiella
26	227.5	10.6	419	2	Q8VQS2	Q8vqs2 klebsiella
27	184	8.5	426	4	Q8BZG2	Q8bzg2 homo sapien
28	133	6.2	416	5	Q19390	Q19390 caenorhabdi
29	132.5	6.2	531	5	Q9GNZ3	Q9gnz3 leishmania
30	123.5	5.7	333	4	Q8ZG3	Q8bzg3 homo sapien
31	123.5	5.7	395	5	Q9VD68	Q9vds8 drosophila
32	123.5	5.7	542	5	Q9UIA2	Q9uia2 leishmania
33	122.5	5.7	683	5	O00838	O00838 leishmania
34	122.5	5.7	707	5	O00839	O00839 leishmania
35	121.5	5.6	447	5	Q9USU3	Q9usu3 drosophila
36	120.5	5.6	447	5	Q9USV1	Q9usv1 drosophila
37	118.5	5.5	442	3	Q8WZJ5	Q8wzj5 aspergillus
38	118.5	5.5	465	3	O00092	O00092 aspergillus
39	117.5	5.5	447	5	Q9USU5	Q9usu5 drosophila
40	117.5	5.5	447	5	Q9TW17	Q9tw17 drosophila
41	117.5	5.5	447	5	Q9TW19	Q9tw19 drosophila
42	117.5	5.5	447	5	Q9U5T5	Q9u5t5 drosophila
43	117.5	5.5	447	5	Q9TW53	Q9tw53 drosophila
44	117.5	5.5	447	5	Q9USU2	Q9usu2 drosophila
45	117.5	5.5	447	5	O97187	O97187 drosophila

ALIGNMENTS

RESULT 1

Q8GN88 PRELIMINARY; PRT; 432 AA.
ID Q8GN88
AC Q8GN88; DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE APPA.
GN APPA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
RN NCBI_TaxID=562;
RX [1]
RP SEQUENCE FROM N.A.
RA Zhang Z., He J., Yao B., Zhou Y., Chen Y., Vi Y.;
RT "Production of phyrase and acid phosphatase by use of silkworm-
bioreactor."
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA Chen Y., Zhu Z., Zhang Z., He J.;
RT "Cloning and overexpression of phyrase gene appA from Escherichia coli."
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
EMBL; AFS37219; AAN28334.1; -
GO; GO:0003993; F:acid phosphatase activity; IEA.
InterPro; IPR000560; HisAc_Phsptase.
Pfam; PF00328; acid phosphat; 1
DR PROSITE; PS00616; HIS-ACID-PHOSPHAT_1; 1.
DR PROSITE; PS00778; HIS-ACID-PHOSPHAT_2; 1.
SQ SEQUENCE 432 AA; 47055 MW; 5B355D76E7377737 CRC64;

Query Match 96.3%; Score 2073; DB 2; Length 432;
Best Local Similarity 97.8%; Pred. No. 2.2e-165;
Matches 401; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

Qy 1 QSEPELKESVIVSRHGVRAPKATQLQMDYTPDAWTPVVKLGELTRGGGLIAYIGH 60
|||||

Db 23 QSEPELKLESVIVSRHGVRAPTKATQMQDVTPDAMPVTPVKLGWLTTPRGGLIAYLGH 82
QY 61 YWRQRLVADGILLPKCCPGSQGQVAIIADVDERTKGTGEAFAAGLAPDCAITVHTQADTSS 120
Db 83 YQORLVADGILLAKKCGPQSGQVAIIADVDERTKGTGEAFAAGLAPDCAITVHTQADTSS 142
QY 121 PDPLFNLKTVGCQOLDNANVTDAILERAGGSIADFTGHYQTAFARELVNFPQSNLCLK 180
Db 143 PDPLFNLKTVGCQOLDNANVTDAILERAGGSIADFTGHYQTAFARELVNFPQSNLCLK 202
QY 181 REKQDESCSLTOALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 240
Db 203 REKQDESCSLTOALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 262
QY 241 HONWTLISLHNAQFDLLQRTPEVARSRATPLLDLIKLTALTPHPQKQAYGVLTPTSVLFI 300
Db 263 HONWTLISLHNAQFDLLQRTPEVARSRATPLLDLIKLTALTPHPQKQAYGVLTPTSVLFI 322
QY 301 AGHDTNLANLGALELNTLPGQDNTPPGSELVFERWRRLSDNSQWISVLFQTLQOM 360
Db 323 AGHDTNLANLGALELNTLPGQDNTPPGSELVFERWRRLSDNSQWISVLFQTLQOM 382
QY 361 RDKTPLSLNTPPGSVKLTLAGCERNAQGMCSLAGFTQIVNEARIPACSL 410
Db 383 RDKTPLSLNTPPGSVKLTLAGCERNAQGMCSLAGFTQIVNEARIPACSL 432

RESULT 2

Q8RKD6 PRELIMINARY; PRT; 432 AA.
AC Q8RKD6; 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, last annotation update)
DE Periplasmic phosphoanhydride phosphohydrolase.
GN APPA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90368616; PubMed=2168385;
RA Dassa J., Marck C., Boquet P.-L.L.;
RT "The complete nucleotide sequence of the Escherichia coli gene appA
RT reveals significant homology between pH 2.5 acid phosphatase and
RT glucose-1-phosphatase.";
RL J. Bacteriol. 172:5497-5500(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93054596; PubMed=1429631;
RA Ostanin K., Harms E.H., Stevis P.E., Kuciel R., Zhou M.M.,
RA Van Etten R.L.;
RT "Overexpression, site-directed mutagenesis, and mechanism of
RT Escherichia coli acid phosphatase.";
RL J. Biol. Chem. 267:22830-22836(1992).
DR EMBL; L03374; AAA00006.1;
DR GO; GO:0003993; F:acid phosphatase activity; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR000560; HisAc_phspatse.
DR Pfam; PF00328; acid_phosphat; 1.
DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
KW Hydrolase.
SQ SEQUENCE 432 AA; 46971 MW; AF86C41EA6193AC5 CRC64;

Query Match 96.2%; Score 2071; DB 2; Length 432;
Best Local Similarity 97.8%; Pred. No. 3.2e-165;
Matches 401; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 1 QSEPELKLESVIVSRHGVRAPTKATQMQDVTPDAMPVTPVKLGWLTTPRGGLIAYLGH 60
Db 23 QSEPELKLESVIVSRHGVRAPTKATQMQDVTPDAMPVTPVKLGWLTTPRGGLIAYLGH 82

QY 61 YWRQRLVADGILLPKCCPGSQGQVAIIADVDERTKGTGEAFAAGLAPDCAITVHTQADTSS 120
Db 83 YQORLVADGILLAKKCGPQSGQVAIIADVDERTKGTGEAFAAGLAPDCAITVHTQADTSS 142
QY 121 PDPLFNLKTVGCQOLDNANVTDAILERAGGSIADFTGHYQTAFARELVNFPQSNLCLK 180
Db 143 PDPLFNLKTVGCQOLDNANVTDAILERAGGSIADFTGHYQTAFARELVNFPQSNLCLK 202
QY 181 REKQDESCSLTOALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 240
Db 203 REKQDESCSLTOALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 262
QY 241 HONWTLISLHNAQFDLLQRTPEVARSRATPLLDLIKLTALTPHPQKQAYGVLTPTSVLFI 300
Db 263 HONWTLISLHNAQFDLLQRTPEVARSRATPLLDLIKLTALTPHPQKQAYGVLTPTSVLFI 322
QY 301 AGHDTNLANLGALELNTLPGQDNTPPGSELVFERWRRLSDNSQWISVLFQTLQOM 360
Db 323 AGHDTNLANLGALELNTLPGQDNTPPGSELVFERWRRLSDNSQWISVLFQTLQOM 382
QY 361 RDKTPLSLNTPPGSVKLTLAGCERNAQGMCSLAGFTQIVNEARIPACSL 410
Db 383 RDKTPLSLNTPPGSVKLTLAGCERNAQGMCSLAGFTQIVNEARIPACSL 432

RESULT 3

Q8RKD7 PRELIMINARY; PRT; 432 AA.
AC Q8RKD7; 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, last annotation update)
DE Periplasmic phosphoanhydride phosphohydrolase.
GN APPA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90368616; PubMed=2168385;
RA Dassa J., Marck C., Boquet P.-L.L.;
RT "The complete nucleotide sequence of the Escherichia coli gene appA
RT reveals significant homology between pH 2.5 acid phosphatase and
RT glucose-1-phosphatase.";
RL J. Bacteriol. 172:5497-5500(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93054596; PubMed=1429631;
RA Ostanin K., Harms E.H., Stevis P.E., Kuciel R., Zhou M.M.,
RA Van Etten R.L.;
RT "Overexpression, site-directed mutagenesis, and mechanism of
RT Escherichia coli acid phosphatase.";
RL J. Biol. Chem. 267:22830-22836(1992).
DR EMBL; L03373; AAA00005.1;
DR GO; GO:0003993; F:acid phosphatase activity; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR000560; HisAc_phspatse.
DR Pfam; PF00328; acid_phosphat; 1.
DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
KW Hydrolase.
SQ SEQUENCE 432 AA; 46971 MW; 755D5E4B1AD916A6 CRC64;

Query Match 96.2%; Score 2071; DB 2; Length 432;
Best Local Similarity 97.8%; Pred. No. 3.2e-165;
Matches 401; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 1 QSEPELKLESVIVSRHGVRAPTKATQMQDVTPDAMPVTPVKLGWLTTPRGGLIAYLGH 60
Db 23 QSEPELKLESVIVSRHGVRAPTKATQMQDVTPDAMPVTPVKLGWLTTPRGGLIAYLGH 82

QY 61 YWQRLVADGLLKKKCPQSGQVAIIADVDERTKTGEAFAAGLAPDCAITVHTQADTSS 120
 DB 83 YQORLVADGLLKKKCPQSGQVAIIADVDERTKTGEAFAAGLAPDCAITVHTQADTSS 142
 QY 121 PDPLFNLKTVGVCQLDNANVTDAIISAGGSIADFTGHYQTAFARELVNFPQSNLCLK 180
 DB 143 PDPLFNLKTVGVCQLDNANVTDAIISAGGSIADFTGHYQTAFARELVNFPQSNLCLK 202
 QY 181 REKQDESCSLTOALPSELKVSADCVSLTGVASLASMTEIFLLQQAQGMPEPGWGRITDS 240
 DB 203 REKQDESCSLTOALPSELKVSADCVSLTGVASLASMTEIFLLQQAQGMPEPGWGRITDS 262
 QY 241 HOWNTLISLHNAQFDLQRTPEVARSRATPLLDLTKTALTPHPQKQAYGVTLPTSVLFI 300
 DB 263 HOWNTLISLHNAQFYLLQRTPEVARSRATPLLDLTKTALTPHPQKQAYGVTLPTSVLFI 322
 QY 301 AGHDTNLANLGGALELNLWTLPGQDNTPPGGELVFERWRRLSDNSQWISLVFQTLQOM 360
 DB 323 AGHDTNLANLGGALELNLWTLPGQDNTPPGGELVFERWRRLSDNSQWISLVFQTLQOM 382
 QY 361 RDKTPLSLNTPPGGVKLTLAGCERNAGCMCSLAGFTQIVNEARIPACSL 410
 DB 383 RDKTPLSLNTPPGGVKLTLAGCERNAGCMCSLAGFTQIVNEARIPACSL 432

RESULT 4
 QSRKEO PRELIMINARY; PRT; 432 AA.
 AC QSRKEO;
 DT 01-JUN-2002 (TREMELrel. 21, Created)
 DT 01-JUN-2002 (TREMELrel. 21, Last sequence update)
 DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
 DE Periplasmic phosphoanhydride phosphohydrolase.
 GN APPA.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90368616; PubMed=2168385;
 RA Dassa J., Marck C., Boquet P.-L.L.;
 RT "The complete nucleotide sequence of the Escherichia coli gene appA
 RT reveals significant homology between pH 2.5 acid phosphatase and
 RT glucose-1-phosphatase.";
 RL J. Bacteriol. 172:5497-5500(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93054596; PubMed=1429631;
 RA Ostanin K., Harms E.H., Stevis P.E., Kuciel R., Zhou M.M.,
 RA Van Etten R.L.;
 RT "Overexpression, site-directed mutagenesis, and mechanism of
 RT Escherichia coli acid phosphatase.";
 RL J. Biol. Chem. 267:22830-22836(1992).
 DR EMBL; L03372; AAA00004.1; -;
 DR GO; GO:0003993; F:acid phosphatase activity; IEA.
 DR GO; GO:0016787; F:hydrolase activity; IEA.
 DR InterPro; IPR000560; HisAc_phsphtse.
 DR Pfam; PF00328; acid_phosphat; 1.
 DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
 KW Hydrolase.
 SQ SEQUENCE 432 AA; 46971 MW; 9A85536B57FCCB5 CRC64;

Query Match 96.2%; Score 2071; DB 2; Length 432;
 Best Local Similarity 97.8%; Pred. No. 3.2e-165;
 Matches 401; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
 QY 1 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTDPAWTTPVKLGELTPRGGLIAYLGH 60
 DB 23 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTDPAWTTPVKLGELTPRGGLIAYLGH 82
 QY 61 YWQRLVADGLLKKKCPQSGQVAIIADVDERTKTGEAFAAGLAPDCAITVHTQADTSS 120

DB 83 YQORLVADGLLKKKCPQSGQVAIIADVDERTKTGEAFAAGLAPDCAITVHTQADTSS 142
 QY 121 PDPLFNLKTVGVCQLDNANVTDAIISAGGSIADFTGHYQTAFARELVNFPQSNLCLK 180
 DB 143 PDPLFNLKTVGVCQLDNANVTDAIISAGGSIADFTGHYQTAFARELVNFPQSNLCLK 202
 QY 181 REKQDESCSLTOALPSELKVSADCVSLTGVASLASMTEIFLLQQAQGMPEPGWGRITDS 240
 DB 203 REKQDESCSLTOALPSELKVSADCVSLTGVASLASMTEIFLLQQAQGMPEPGWGRITDS 262
 QY 241 HOWNTLISLHNAQFDLQRTPEVARSRATPLLDLTKTALTPHPQKQAYGVTLPTSVLFI 300
 DB 263 HOWNTLISLHNAQFYLLQRTPEVARSRATPLLDLTKTALTPHPQKQAYGVTLPTSVLFI 322
 QY 301 AGHDTNLANLGGALELNLWTLPGQDNTPPGGELVFERWRRLSDNSQWISLVFQTLQOM 360
 DB 323 AGHDTNLANLGGALELNLWTLPGQDNTPPGGELVFERWRRLSDNSQWISLVFQTLQOM 382
 QY 361 RDKTPLSLNTPPGGVKLTLAGCERNAGCMCSLAGFTQIVNEARIPACSL 410
 DB 383 RDKTPLSLNTPPGGVKLTLAGCERNAGCMCSLAGFTQIVNEARIPACSL 432

RESULT 5
 QSRKEO PRELIMINARY; PRT; 432 AA.
 AC QSRKEO;
 DT 01-JUN-2002 (TREMELrel. 21, Created)
 DT 01-JUN-2002 (TREMELrel. 21, Last sequence update)
 DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
 DE Periplasmic phosphoanhydride phosphohydrolase.
 GN APPA.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90368616; PubMed=2168385;
 RA Dassa J., Marck C., Boquet P.-L.L.;
 RT "The complete nucleotide sequence of the Escherichia coli gene appA
 RT reveals significant homology between pH 2.5 acid phosphatase and
 RT glucose-1-phosphatase.";
 RL J. Bacteriol. 172:5497-5500(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93054596; PubMed=1429631;
 RA Ostanin K., Harms E.H., Stevis P.E., Kuciel R., Zhou M.M.,
 RA Van Etten R.L.;
 RT "Overexpression, site-directed mutagenesis, and mechanism of
 RT Escherichia coli acid phosphatase.";
 RL J. Biol. Chem. 267:22830-22836(1992).
 DR EMBL; L03370; AAA00002.1; -;
 DR GO; GO:0003993; F:acid phosphatase activity; IEA.
 DR GO; GO:0016787; F:hydrolase activity; IEA.
 DR InterPro; IPR000560; HisAc_phsphtse.
 DR Pfam; PF00328; acid_phosphat; 1.
 DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
 KW Hydrolase.
 SQ SEQUENCE 432 AA; 46971 MW; 5BBD632D4682EABF CRC64;

Query Match 96.2%; Score 2071; DB 2; Length 432;
 Best Local Similarity 97.8%; Pred. No. 3.2e-165;
 Matches 401; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
 QY 1 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTDPAWTTPVKLGELTPRGGLIAYLGH 60
 DB 23 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTDPAWTTPVKLGELTPRGGLIAYLGH 82
 QY 61 YWQRLVADGLLKKKCPQSGQVAIIADVDERTKTGEAFAAGLAPDCAITVHTQADTSS 120
 DB 83 YQORLVADGLLKKKCPQSGQVAIIADVDERTKTGEAFAAGLAPDCAITVHTQADTSS 142

QY 121 PDPLFNLKTVGCQOLDNANVTDAILRAGGSIADFTGHTQTAFARELERVLPFQSNLCLK 180
 DB 143 PDPLFNLKTVGCQOLDNANVTDAILRAGGSIADFTGHTQTAFARELERVLPFQSNLCLK 202
 QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGWPEPGWGRITDS 240
 DB 203 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGWPEPGWGRITDS 262
 QY 241 HONWTLISLHNAQFDLLQRTPEVARSRATPLDLLIKTALTTPHPKQAYGVTLPTSVLFI 300
 DB 263 HONWTLISLHNAQFYLLQRTPEVARSRATPLDLLIKTALTTPHPKQAYGVTLPTSVLFI 322
 QY 301 AGHDTNLANLGGALELNLWTLPGQDNTPPGGELVFERWRRLSDNSQWIVQVSLVFQTLQOM 360
 DB 323 AGHDTNLANLGGALELNLWTLPGQDNTPPGGELVFERWRRLSDNSQWIVQVSLVFQTLQOM 382
 QY 361 RDKTPLSLNTPPGEVKLTLAGCEERNAQMCISLAGFTQIVNEARIPACSL 410
 DB 383 RDKTPLSLNTPPGEVKLTLAGCEERNAQMCISLAGFTQIVNEARIPACSL 432

RESULT 6

QSRKD9 PRELIMINARY; PRT; 432 AA.
 ID QSRKD9
 AC QSRKD9;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Periplasmic phosphoanhydride phosphohydrolase.
 GN APPA.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90368616; PubMed=2169385;
 RA Dassa J., Marck C., Boquet P.-L.L.;
 RT "The complete nucleotide sequence of the Escherichia coli gene appA
 RT reveals significant homology between pH 2.5 acid phosphatase and
 RT glucose-1-phosphatase."
 RL J. Bacteriol. 172:5497-5500 (1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93054596; PubMed=1429631;
 RA Ostanin K., Harms E.H., Stevis P.E., Kuciel R., Zhou M.M.,
 RA Van Erten R.L.;
 RT "Overexpression, site-directed mutagenesis, and mechanism of
 RT Escherichia coli acid phosphatase."
 RL J. Biol. Chem. 267:22830-22836 (1992).
 DR EMBL; L03371; AAA00003.1; -
 DR GO; GO:0003993; F:acid phosphatase activity; IEA.
 DR GO; GO:0016787; F:hydrolase activity; IEA.
 DR InterPro; IPR000560; HisAc phsphtse.
 DR Pfam; PF00328; acid_phosphat; 1.
 DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
 KW Hydrolase.
 SQ SEQUENCE 432 AA; 47033 MW; 9F2959DF9C368175 CRC64;

Query Match 96.1%; Score 2070; DB 2; Length 432;
 Best Local Similarity 97.8%; Pred. No. 3.9e-165;
 Matches 401; Conservative 1; Mismatches 8; Indels 0; Gaps 0;
 QY 1 QSEPELKESVVIYSRHGVRAPTKATQMDVTPDAMPVFWKLGELTPRGGLIAYLGH 60
 DB 23 QSEPELKESVVIYSRHGVRAPTKATQMDVTPDAMPVFWKLGELTPRGGLIAYLGH 82
 QY 61 YWRQLVADGLLKKCGCPQSGQVAIIADVDERTKTGEAFAGLAPDCAITVHTQADTSS 120
 DB 83 YWRQLVADGLLKKCGCPQSGQVAIIADVDERTKTGEAFAGLAPDCAITVHTQADTSS 142
 QY 121 PDPLFNLKTVGCQOLDNANVTDAILRAGGSIADFTGHTQTAFARELERVLPFQSNLCLK 180

DB 143 PDPLFNLKTVGCQOLDNANVTDAILRAGGSIADFTGHTQTAFARELERVLPFQSNLCLK 202
 QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGWPEPGWGRITDS 240
 DB 203 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGWPEPGWGRITDS 262
 QY 241 HONWTLISLHNAQFDLLQRTPEVARSRATPLDLLIKTALTTPHPKQAYGVTLPTSVLFI 300
 DB 263 HONWTLISLHNAQFYLLQRTPEVARSRATPLDLLIKTALTTPHPKQAYGVTLPTSVLFI 322
 QY 301 AGHDTNLANLGGALELNLWTLPGQDNTPPGGELVFERWRRLSDNSQWIVQVSLVFQTLQOM 360
 DB 323 AGHDTNLANLGGALELNLWTLPGQDNTPPGGELVFERWRRLSDNSQWIVQVSLVFQTLQOM 382
 QY 361 RDKTPLSLNTPPGEVKLTLAGCEERNAQMCISLAGFTQIVNEARIPACSL 410
 DB 383 RDKTPLSLNTPPGEVKLTLAGCEERNAQMCISLAGFTQIVNEARIPACSL 432

RESULT 7

QSRKD5 PRELIMINARY; PRT; 432 AA.
 ID QSRKD5;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Periplasmic phosphoanhydride phosphohydrolase.
 GN APPA.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90368616; PubMed=2169385;
 RA Dassa J., Marck C., Boquet P.-L.L.;
 RT "The complete nucleotide sequence of the Escherichia coli gene appA
 RT reveals significant homology between pH 2.5 acid phosphatase and
 RT glucose-1-phosphatase."
 RL J. Bacteriol. 172:5497-5500 (1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93054596; PubMed=1429631;
 RA Ostanin K., Harms E.H., Stevis P.E., Kuciel R., Zhou M.M.,
 RA Van Erten R.L.;
 RT "Overexpression, site-directed mutagenesis, and mechanism of
 RT Escherichia coli acid phosphatase."
 RL J. Biol. Chem. 267:22830-22836 (1992).
 DR EMBL; L03371; AAA00003.1; -
 DR GO; GO:0003993; F:acid phosphatase activity; IEA.
 DR GO; GO:0016787; F:hydrolase activity; IEA.
 DR InterPro; IPR000560; HisAc phsphtse.
 DR Pfam; PF00328; acid_phosphat; 1.
 DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
 KW Hydrolase.
 SQ SEQUENCE 432 AA; 46990 MW; 951F393EA9A1A47C CRC64;

Query Match 96.0%; Score 2067; DB 2; Length 432;
 Best Local Similarity 97.8%; Pred. No. 6.9e-165;
 Matches 401; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
 QY 1 QSEPELKESVVIYSRHGVRAPTKATQMDVTPDAMPVFWKLGELTPRGGLIAYLGH 60
 DB 23 QSEPELKESVVIYSRHGVRAPTKATQMDVTPDAMPVFWKLGELTPRGGLIAYLGH 82
 QY 61 YWRQLVADGLLKKCGCPQSGQVAIIADVDERTKTGEAFAGLAPDCAITVHTQADTSS 120
 DB 83 YWRQLVADGLLKKCGCPQSGQVAIIADVDERTKTGEAFAGLAPDCAITVHTQADTSS 142
 QY 121 PDPLFNLKTVGCQOLDNANVTDAILRAGGSIADFTGHTQTAFARELERVLPFQSNLCLK 180
 DB 143 PDPLFNLKTVGCQOLDNANVTDAILRAGGSIADFTGHTQTAFARELERVLPFQSNLCLK 202

```
QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMGPEPGWGRITDS 240
Db 203 REKQDESCSLTQALPSELKVSADNVSLTGAVSLASMLTEIFLLQQAQMGPEPGWGRITDS 262
QY 241 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLDLIKTALTALTPHPKQAYGVTLPTSVLFI 300
Db 263 HQWNTLLSLHNAQFYLLQRTPEVARSRATPLDLIKTALTALTPHPKQAYGVTLPTSVLFI 322
QY 301 AGHDTNLANLGGALELNTLPGQPDNTPGCGELVFERWRRLSDNSQWISLVFQTLQOM 360
Db 323 AGADTNLANLGGALELNTLPGQPDNTPGCGELVFERWRRLSDNSQWISLVFQTLQOM 382
QY 361 RDKTPLSLNTPPGGVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 410
Db 383 RDKTPLSLNTPPGGVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 432

RESULT 8
Q83RW2 PRELIMINARY; PRT; 442 AA.
ID AC Q83RW2
AC Q83RW2;
DT 01-JUN-2003 (TREMREL. 24, Created)
DT 01-JUN-2003 (TREMREL. 24, Last sequence update)
DT 01-OCT-2003 (TREMREL. 25, Last annotation update)
DE Phosphoanhydride phosphorylase, pH 2.5 acid phosphatase,
DE periplasmic.
GN APPA OR SF0982.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=301 / Serotype 2a;
RX MEDLINE=22272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qi D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Chang H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157."
RL Nucleic Acids Res. 30:4432-4441(2002).
DR EMBL; AE015127; AAN2610.1; -
DR GO; GO:0003993; P:acid phosphatase activity; IEA.
DR InterPro; IPR000560; HisAc_phsphtse.
DR Pfam; PF00328; acid_phosphat; 1.
DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
KW Complete proteome.
SQ SEQUENCE 442 AA; 48183 MW; 20BCE2C454254680 CRC64;

Query Match 94.9%; Score 2044; DB 16; Length 442;
Best Local Similarity 96.6%; Pred. No. 6.1e-163;
Matches 396; Conservative 1; Mismatches 13; Indels 0; Gaps 0;

QY 1 QSEPELKLESVIVSRHGVRAPTKATQLMQDVTDPDWTVPVKLGELTPRGGLIAYLGH 60
Db 33 QSEPELKLESVIVSRHGVRAPTKATQLMQDVTDPDWTVPVKLGELTPRGGLIAYLGH 92
QY 61 YWRQLVADGLLPKCGCPQSGQVAIIADVDERTRKTGEAFAGLAPDCAITVHTQADTSS 120
Db 93 YWRQLVADGLLPKCGCPQSGQVAIIADVDERTRKTGEAFAGLAPDCAITVHTQADTSS 152
QY 121 PDPLFNPLKTVGCQLDNANVTDAILLERAGGSFADPTGHVQTAFRELERVLPFQSNCLCK 180
Db 153 PDPLFNPLKTVGCQLDNANVTDAILLERAGGSFADPTGHVQTAFRELERVLPFQSNCLCK 212
QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMGPEPGWGRITDS 240
Db 213 REKQDESCSLTQALPSELKVSADNVSLTGAVSLASMLTEIFLLQQAQMGPEPGWGRITDS 272
QY 241 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLDLIKTALTALTPHPKQAYGVTLPTSVLFI 300
```

```
Db 273 HQWNTLLSLHNAQFYLLQRTPEVARSRATPLDLIKTALTALTPHPKQAYGVTLPTSVLFI 332
QY 301 AGHDTNLANLGGALELNTLPGQPDNTPGCGELVFERWRRLSDNSQWISLVFQTLQOM 360
Db 333 AGHDTNLANLGGALELNTLPGQPDNTPGCGELVFERWRRLSDNSQWISLVFQTLQOM 392
QY 361 RDKTPLSLNTPPGGVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 410
Db 393 RDKTPLSLNTPPGGVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 442

RESULT 9
Q8XC29 PRELIMINARY; PRT; 434 AA.
ID AC Q8XC29
AC Q8XC29;
DT 01-MAR-2002 (TREMREL. 20, Created)
DT 01-MAR-2002 (TREMREL. 20, Last sequence update)
DT 01-OCT-2003 (TREMREL. 25, Last annotation update)
DE Phosphoanhydride phosphorylase, pH 2.5 acid phosphatase,
DE periplasmic.
GN APPA OR Z1397 OR ECS1136.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Link S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Kihara S., Shiba H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
DR EMBL; AE005292; AAG55528.1; ALT INIT.
DR EMBL; AP002554; BAB34559.1; -
DR PIR; D85633; D85633.
DR PIR; H90770; H90770.
DR GO; GO:0003993; P:acid phosphatase activity; IEA.
DR InterPro; IPR000560; HisAc_phsphtse.
DR Pfam; PF00328; acid_phosphat; 1.
DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
KW Complete proteome.
SQ SEQUENCE 434 AA; 47337 MW; F197DF7D1869F9C4 CRC64;

Query Match 94.8%; Score 2042; DB 16; Length 434;
Best Local Similarity 96.3%; Pred. No. 8.7e-163;
Matches 395; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

QY 1 QSEPELKLESVIVSRHGVRAPTKATQLMQDVTDPDWTVPVKLGELTPRGGLIAYLGH 60
Db 25 EPEPELKLESVIVSRHGVRAPTKATQLMQDVTDPDWTVPVKLGELTPRGGLIAYLGH 84
QY 61 YWRQLVADGLLPKCGCPQSGQVAIIADVDERTRKTGEAFAGLAPDCAITVHTQADTSS 120
Db 85 YWRQLVADGLLPKCGCPQSGQVAIIADVDERTRKTGEAFAGLAPDCAITVHTQADTSS 144
```



```

QY 121 PDPLFNPLKTVGCQDQDANVTDAIILRAGGSIADFTGHYQTAFRELERVLPQSNLCLK 180
DB 145 PDPLFNPLKTVGCQDQDANVTDAIILRAGGSIADFTGHYQTAFRELERVLPQSNLCLK 204
QY 181 REKQDESCSLTOALPSELKVSADCVSLTGAVSLASMLTEIFLLQOAGMPGCGRITDS 240
DB 205 REKQDESCSLTOALPSELKVSADCVSLTGAVSLASMLTEIFLLQOAGMPGCGRITDS 264
QY 241 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLLDLIKTAITPHPPQKQAYGVTLPSTVLF 300
DB 265 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLLDLIKTAITPHPPQKQAYGVTLPSTVLF 324
QY 301 AGHDTNLANLGGALELNWTLPGQDNTPPGGELVFERWRRLSDNSOWIQVSLVFQTLQOM 360
DB 325 AGHDTNLANLGGALELNWTLPGQDNTPPGGELVFERWRRLSDNSOWIQVSLVFQTLQOM 384
QY 361 RDKTPLSLNTPPGVVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 410
DB 385 RDKTPLSLNTPPGVVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 434
RESULT 10
Q8CW75 PRELIMINARY; PRT; 446 AA.
AC Q8CW75;
DT 01-WAR-2003 (TremBLrel. 23, Created)
DT 01-WAR-2003 (TremBLrel. 23, Last sequence update)
DE Periplasmic appA protein precursor.
GN APPA OR C1121.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Raeko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.; Perna N.T.,
RT "Extensive mosaic structure revealed by the complete genome sequence
of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
DR EMBL; AE016759; AAN79589.1; -.
GO: GO:0003993; F:acid phosphatase activity; IEA.
DR InterPro; IPR000560; HisAcPhsphtase.
DR Pfam; PF00328; acid_phosphat; 1.
DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
KW Complete proteome.
SQ SEQUENCE 446 AA; 48605 MW; F1308CD691DB9F49 CRC64;
Query Match 94.8%; Score 2042; DB 16; Length 446;
Best Local Similarity 96.3%; Pred. No. 9.1e-163;
Matches 395; Conservative 2; Mismatches 13; Indels 0; Gaps 0;
QY 1 QSEPELKESVIVSRHGVRAFTKATQMQDVTDPDAMPTWPKLGELTPRGGLIAYLGH 60
DB 33 QSEPELKESVIVSRHGVRAFTKATQMQDVTDPDAMPTWPKLGELTPRGGLIAYLGH 92
QY 61 YRQRLVADGLLPKCGCPSQSGOVAIIADVDERTKTGEAFAGLAPDCAITVHTQADTSS 120
DB 93 YRQRLVADGLLPKCGCPSQSGOVAIIADVDERTKTGEAFAGLAPDCAITVHTQADTSS 152
QY 121 PDPLFNPLKTVGCQDQDANVTDAIILRAGGSIADFTGHYQTAFRELERVLPQSNLCLK 180
DB 153 PDPLFNPLKTVGCQDQDANVTDAIILRAGGSIADFTGHYQTAFRELERVLPQSNLCLK 212
QY 181 REKQDESCSLTOALPSELKVSADCVSLTGAVSLASMLTEIFLLQOAGMPGCGRITDS 240
DB 213 REKQDESCSLTOALPSELKVSADCVSLTGAVSLASMLTEIFLLQOAGMPGCGRITDS 272
Query Match 94.8%; Score 2042; DB 16; Length 446;
Best Local Similarity 96.3%; Pred. No. 9.1e-163;
Matches 395; Conservative 2; Mismatches 13; Indels 0; Gaps 0;
QY 1 QSEPELKESVIVSRHGVRAFTKATQMQDVTDPDAMPTWPKLGELTPRGGLIAYLGH 60
DB 33 QSEPELKESVIVSRHGVRAFTKATQMQDVTDPDAMPTWPKLGELTPRGGLIAYLGH 92
QY 61 YRQRLVADGLLPKCGCPSQSGOVAIIADVDERTKTGEAFAGLAPDCAITVHTQADTSS 120
DB 93 YRQRLVADGLLPKCGCPSQSGOVAIIADVDERTKTGEAFAGLAPDCAITVHTQADTSS 152
QY 121 PDPLFNPLKTVGCQDQDANVTDAIILRAGGSIADFTGHYQTAFRELERVLPQSNLCLK 180
DB 153 PDPLFNPLKTVGCQDQDANVTDAIILRAGGSIADFTGHYQTAFRELERVLPQSNLCLK 212
QY 181 REKQDESCSLTOALPSELKVSADCVSLTGAVSLASMLTEIFLLQOAGMPGCGRITDS 240
DB 213 REKQDESCSLTOALPSELKVSADCVSLTGAVSLASMLTEIFLLQOAGMPGCGRITDS 272

```

```

QY 241 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLLDLIKTAITPHPPQKQAYGVTLPSTVLF 300
DB 273 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLLDLIKTAITPHPPQKQAYGVTLPSTVLF 332
QY 301 AGHDTNLANLGGALELNWTLPGQDNTPPGGELVFERWRRLSDNSOWIQVSLVFQTLQOM 360
DB 333 AGHDTNLANLGGALELNWTLPGQDNTPPGGELVFERWRRLSDNSOWIQVSLVFQTLQOM 392
QY 361 RDKTPLSLNTPPGVVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 410
DB 393 RDKTPLSLNTPPGVVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 442
RESULT 11
Q7UD08 PRELIMINARY; PRT; 432 AA.
AC Q7UD08;
DT 01-OCT-2003 (TremBLrel. 25, Created)
DT 01-OCT-2003 (TremBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE Phosphoanhydride phosphorylase; pH 2.5 acid phosphatase.
GN APPA OR S1048.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=22590274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyane-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
flexneri serotype 2a strain 2457T.";
RL Infect. Immun. 71:2775-2786 (2003).
DR EMBL; AE016981; AAP16495.1; -.
SQ SEQUENCE 432 AA; 47136 MW; 200442239BC66DFC CRC64;
Query Match 94.8%; Score 2040; DB 16; Length 432;
Best Local Similarity 96.3%; Pred. No. 1.3e-162;
Matches 395; Conservative 1; Mismatches 14; Indels 0; Gaps 0;
QY 1 QSEPELKESVIVSRHGVRAFTKATQMQDVTDPDAMPTWPKLGELTPRGGLIAYLGH 60
DB 23 QSEPELKESVIVSRHGVRAFTKATQMQDVTDPDAMPTWPKLGELTPRGGLIAYLGH 82
QY 61 YRQRLVADGLLPKCGCPSQSGOVAIIADVDERTKTGEAFAGLAPDCAITVHTQADTSS 120
DB 83 YRQRLVADGLLPKCGCPSQSGOVAIIADVDERTKTGEAFAGLAPDCAITVHTQADTSS 142
QY 121 PDPLFNPLKTVGCQDQDANVTDAIILRAGGSIADFTGHYQTAFRELERVLPQSNLCLK 180
DB 143 PDPLFNPLKTVGCQDQDANVTDAIILRAGGSIADFTGHYQTAFRELERVLPQSNLCLK 202
QY 181 REKQDESCSLTOALPSELKVSADCVSLTGAVSLASMLTEIFLLQOAGMPGCGRITDS 240
DB 203 REKQDESCSLTOALPSELKVSADCVSLTGAVSLASMLTEIFLLQOAGMPGCGRITDS 262
QY 241 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLLDLIKTAITPHPPQKQAYGVTLPSTVLF 300
DB 263 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLLDLIKTAITPHPPQKQAYGVTLPSTVLF 322
QY 301 AGHDTNLANLGGALELNWTLPGQDNTPPGGELVFERWRRLSDNSOWIQVSLVFQTLQOM 360
DB 323 AGHDTNLANLGGALELNWTLPGQDNTPPGGELVFERWRRLSDNSOWIQVSLVFQTLQOM 382
QY 361 RDKTPLSLNTPPGVVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 410
DB 383 RDKTPLSLNTPPGVVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 432

```

```
RESULT 12
Q8ZFP6 PRELIMINARY; PRT; 441 AA.
AC Q8ZFP6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Probable histidine acid phosphatase (EC 3.1.3.2) (Phosphoanhydride
phosphorylase).
GN YP01648 OR APPA OR Y1810.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Fellwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moulé S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;
RA "Genome sequence of Yersinia pestis, the causative agent of plague.";
RT Nature 413:523-527(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=KIM5 / Biovar Mediaevalis;
RX MEDLINE=22137863; PubMed=12142430;
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA Fetherston J.D., Lindler L.E., Brubaker R.W., Plano G.V.,
RA Straley S.C., McDonough K.A., Nilles M.D., Matson J.S., Blattner F.R.,
RA Perry R.D.;
RT "Genome sequence of Yersinia pestis KIM.";
RN J. Bacteriol. 184:4601-4611(2002).
DR EMBL; AJ414149; CAC90470.1; --
DR EMBL; AE013783; AAM85378.1; --
DR PIR; AC0201; AC0201.
DR GO; GO:0003993; F:acid phosphatase activity; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR000560; HisAc_phsphtase.
DR Pfam; PF00328; acid_phosphat_1.
DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR Hydrolase; Complete proteome.
KW SEQUENCE 441 AA; 47784 MW; CDD15COA64C98439 CRC64;

Query Match 44.3%; Score 954; DB 16; Length 441;
Best Local Similarity 46.8%; Pred. No. 1.7e-71;
Matches 195; Conservative 65; Mismatches 145; Indels 12; Gaps 6;

QY 2 SEPE-LKLESVVVSRHGVRAP-TKATQLMQDVTPDAMPVTKVKGELTPGGELIAYLGH 60
DB 29 AEPGTYLLEVRVILSRHGVRSPTKQTQLMNDVTPDKWPQMPVKAGYLTTPGAEILVTLMGG 88
QY 61 YWRQLVADGLLPKCGCPQSQGVVAIADVDERTKTKGEAFAGLAPDCAITVHTQADTSS 120
DB 89 FYGDFYRSLGLL-AAGCPAEGGVVAQADIDQRTLTGQFLDGVAPCGGLTVHQNADLKK 147
QY 121 PDPLFNPLKTVGCQOLDNANVTDAILERAGGSIAFTGHYQTAFRELRLVNFPSNCLK 180
DB 148 TDPLFHPVEAGVCKLDAQAOTDKAIEEQGLGSLDVTVSQRYAKPPAQMGDVLNFAASPYCKS 207
QY 181 REKQDESCLTAQALPSELKVSAD--CVSLTCAVSLASMLTIFLLQQAQMGPEPGWRIT 238
DB 208 LQQQKTCDFAFHFAENVNKEGTVKLTSLSPALSSLTGIFLLQNAQMPVAVQRLX 267
QY 239 DSHQNTLLSLHNAQFDLLQRTPEVARSRAFLDLIKTALTTPPPQKQAVGVLTPTS-- 296
DB 268 GAENWVSLSLHNAQFLNMAKTPYIARHKGTFLQLQIDTALT---LQIDAQGGQLPISAQ 324

RESULT 13
Q8XBZ6 PRELIMINARY; PRT; 413 AA.
AC Q8XBZ6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Periplasmic glucose-1-phosphatase.
GN AGP OR Z1421 OR ECS1158.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoudis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RN Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayaishi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tohe T.,
RA Iida T., Takami H., Honda Y., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12.";
RN DNA Res. 8:11-22(2001).
DR EMBL; AE005294; AAG55550.1; --
DR EMBL; AP002554; BAB34581.1; --
DR PIR; B85636; B85636.
DR PIR; F90773; F90773.
DR GO; GO:0003993; F:acid phosphatase activity; IEA.
DR InterPro; IPR000560; HisAc_phsphtase.
DR Pfam; PF00328; acid_phosphat_1.
DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
KW Complete proteome.
QY SEQUENCE 413 AA; 45640 MW; A4630644EF6FE1A CRC64;

Query Match 25.6%; Score 551; DB 16; Length 413;
Best Local Similarity 33.3%; Pred. No. 1e-37;
Matches 138; Conservative 71; Mismatches 169; Indels 36; Gaps 12;

QY 1 QSEPE-LKLESVVVSRHGVRAP-TKATQLMQDVTPDAMPVTKVKGELTPGGELIAYL 58
DB 23 QTVPEGYQQVQLVMSRRHLRAPLANNQSVLEQSTFNKMPEDVPGQJTTKGGVLEVYM 82
QY 59 GHYWRQLVADGLLPKCGCPQSQGVVAIADVDERTKTKGEAFAGLAPDCAITVHTQADT 118
DB 83 GHYWRQLVADGLLPKCGCPQSQGVVAIADVDERTKTKGEAFAGLAPDCAITVHTQADT 142
QY 119 SPDPFLFNPLKTVGCQOLDNANVTDAILERAGGSIAFTGHYQTAFRELRLVNFPS 175
DB 143 GTMDPTFNFVITD---DSAAFSEQAVAMEKLSKL-QLTDSYQL---LEKIVNYKDS 193
```


Db	250	WGBIKSPQWKLKNGYQDSLFTSPEVARNVAKPLVSYIDKALV--TDRASAPKITV	307
Qy	294	PTSVLFIAGHDTNLANIGGAELELW-LPGQPDNTPPGGELVFERWRRLSDNSQWIOVSL	352
Db	308	-----LVGHDSNIALTLTALDFKPYQLHDQNEBTPIGGKIVFORWHDSCANRDLMKIEY	361
Qy	353	VFQLOQMRDKTPISLNTPPEGEVKLTLAGCEBERNAQCMCSLAGFTQIVNEA	403
Db	362	VYQSAEQLRNADALTLOAPAQRVTLLESGC-PIDANGFCFMDKFDVSVLNEA	411

Search completed: April 30, 2004, 12:41:25
Job time : 38.1429 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 30, 2004, 12:41:31 ; Search time 40 Seconds
(without alignments)
2841.193 Million cell updates/sec

Title: US-10-021-723B-13
 Perfect score: 2153
 Sequence: 1 OSEPDKUESWIVSHGVR.....CSLAGTQIVNEARIPACSL 410

Scoring table: BLOSUM62

Searched: 1138120 segs, 277189581 residues

Total number of hits satisfying chosen parameters: 1138120

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Published Applications AA:*

1: /cgm2_6/ptodata/1/pubpaa/US07_PUBCOMB.pcp:
2: /cgm2_6/ptodata/1/pubpaa/US07_NEW_PUB.pcp:
3: /cgm2_6/ptodata/1/pubpaa/US06_NEW_PUB.pcp:
4: /cgm2_6/ptodata/1/pubpaa/US06_PUBCOMB.pcp:
5: /cgm2_6/ptodata/1/pubpaa/US07_NEW_PUB.pcp:
6: /cgm2_6/ptodata/1/pubpaa/US07_PUBCOMB.pcp:
7: /cgm2_6/ptodata/1/pubpaa/US08_NEW_PUB.pcp:
8: /cgm2_6/ptodata/1/pubpaa/US08_PUBCOMB.pcp:
9: /cgm2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pcp:
10: /cgm2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pcp:
11: /cgm2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pcp:
12: /cgm2_6/ptodata/1/pubpaa/US09_NEW_PUB.pcp:
13: /cgm2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pcp:
14: /cgm2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pcp:
15: /cgm2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pcp:
16: /cgm2_6/ptodata/1/pubpaa/US10_NEW_PUB.pcp:
17: /cgm2_6/ptodata/1/pubpaa/US60_NEW_PUB.pcp:
18: /cgm2_6/ptodata/1/pubpaa/US60_PUBCOMB.pcp:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query #		Length	DB	ID	Description
		Match					
1	2153	100.0	410	14	US-10-021-723A-13		Sequence 13, Appl
2	2153	100.0	410	14	US-10-021-723A-15		Sequence 15, Appl
3	2153	100.0	412	14	US-10-334-672-1		Sequence 1, Appl
4	2153	100.0	412	14	US-10-334-671-1		Sequence 1, Appl
5	2153	100.0	436	14	US-10-156-660-2		Sequence 2, Appl
6	2123	98.6	430	9	US-09-866-379-10		Sequence 10, Appl
7	2077	96.5	432	9	US-09-866-378-8		Sequence 8, Appl
8	2077	96.5	432	12	US-10-282-122A-43351		Sequence 43351, A
9	2077	96.5	432	14	US-10-156-660-4		Sequence 4, Appl
10	2066	96.0	440	9	US-09-777-566A-2		Sequence 2, Appl
11	2066	96.0	440	9	US-09-866-379-2		Sequence 2, Appl
12	2066	96.0	440	14	US-10-034-385-2		Sequence 2, Appl
13	2066	96.0	440	15	US-10-430-356-2		Sequence 2, Appl
14	2042	94.8	432	15	US-10-284-962-3		Sequence 3, Appl
15	2042	94.8	432	15	US-10-284-962-14		Sequence 14, Appl

16	2042	94.8	433	14	US-10-266-041-1	Sequence 1, Appl
17	2039	94.7	432	15	US-10-284-962-5	Sequence 5, Appl
18	2007	93.2	432	14	US-10-021-723A-16	Sequence 16, Appl
19	971	45.1	441	14	US-10-021-723A-12	Sequence 12, Appl
20	954	44.3	441	12	US-10-282-123A-77792	Sequence 77792, A
21	946	43.9	441	14	US-10-021-723A-2	Sequence 2, Appl
22	943	43.8	420	14	US-10-021-723A-4	Sequence 4, Appl
23	773.5	35.9	476	14	US-10-021-723A-10	Sequence 10, Appl
24	613.5	28.5	318	14	US-10-021-723A-14	Sequence 14, Appl
25	483.5	21.5	409	14	US-10-021-723A-8	Sequence 8, Appl
26	458.5	21.3	421	14	US-10-021-723A-6	Sequence 6, Appl
27	184	8.5	426	12	US-10-257-174-44	Sequence 44, Appl
28	184	8.5	426	12	US-10-343-357-6	Sequence 6, Appl
29	132.5	5.7	99	12	US-10-282-123A-59099	Sequence 59099, A
30	119.5	5.6	467	12	US-10-442-538-165	Sequence 165, A
31	118.5	5.5	439	14	US-10-062-848-3	Sequence 3, Appl
32	118.5	5.5	440	12	US-10-442-538-106	Sequence 106, App
33	118.5	5.5	440	12	US-10-442-538-128	Sequence 128, App
34	118.5	5.5	440	12	US-10-442-538-151	Sequence 151, App
35	118.5	5.5	449	14	US-10-062-848-12	Sequence 12, Appl
36	118.5	5.5	455	13	US-10-083-452-8	Sequence 8, Appl
37	118.5	5.5	465	14	US-10-062-848-78	Sequence 78, Appl
38	118.5	5.5	465	14	US-10-229-358-6	Sequence 6, Appl
39	118.5	5.5	474	14	US-10-213-990-24	Sequence 24, Appl
40	117.5	5.5	440	12	US-10-442-538-107	Sequence 107, App
41	117.5	5.5	440	12	US-10-442-538-129	Sequence 129, App
42	117.5	5.5	440	12	US-10-442-538-152	Sequence 152, App
43	110.5	5.1	440	12	US-10-442-538-108	Sequence 108, App
44	110.5	5.1	440	12	US-10-442-538-130	Sequence 130, App
45	110.5	5.1	440	12	US-10-442-538-153	Sequence 153, App

ALIGNMENTS

```

RESULT 1
US-10-021-723A-13
; Sequence 13, Application US/10021723A
; Publication NO. US20030101476A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Mathur, Eric
; APPLICANT: Richardson, Toby
; APPLICANT: Robertson, Dan
; APPLICANT: Barton, Nelson
; TITLE OF INVENTION: Recombinant Phytases and Uses Thereof
; FILE REFERENCE: 113766.140 (DIV-016CIP)
; CURRENT APPLICATION NUMBER: US/10/021,723A
; CURRENT FILING DATE: 2002-10-23
; PRIOR APPLICATION NUMBER: US 607255,090
; PRIOR FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 410
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: amino acid sequence of a phytase enzyme
; OTHER INFORMATION: predicted glycosylation sites
US-10-021-723A-13

```

	Query Match	Best Local Similarity	Mismatches	Indels	Gaps	Length
	100.0%;	100.0%;	0;	0;	0;	410;
	Score 2153;	Fred. No. 2.3e-193;	Mismatches 0;	Indels 0;	Gaps 0;	Length 410;
QY	1	QSEPELKLSVIVSRHGVRAPTKATQLMDVTPDAMPVKLGELTPRGGLIAYLGH	60			
Db	1	QSEPELKLSVIVSRHGVRAPTKATQLMDVTPDAMPVKLGELTPRGGLIAYLGH	60			
QY	61	YVORIVADGLLPKCSCPSGOVAILADYDERTKTGEAAGLAPDCAITVTDQTSS	120			
Db	61	YVORIVADGLLPKCSCPSGOVAILADYDERTKTGEAAGLAPDCAITVTDQTSS	120			

QY 121 PDPLFNPLKTVGCOLDNANVTDAILERAGGSIADFTGHYQTAFARELERVLPQSNLCLK 180
 Db 121 PDPLFNPLKTVGCOLDNANVTDAILERAGGSIADFTGHYQTAFARELERVLPQSNLCLK 180
 QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 240
 Db 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 240
 QY 241 HOWNTLLSLHNAQFDLLQRTPEVARSRATPLDLIKTALTALTPHPPOKQAYGVTLPSTVLF 300
 Db 241 HOWNTLLSLHNAQFDLLQRTPEVARSRATPLDLIKTALTALTPHPPOKQAYGVTLPSTVLF 300
 QY 301 AGHDTNLANLGGALELAWNTLPGQDNTPPGGELVFERWRRLSDNSQWIOVSLVFTQLQOM 360
 Db 301 AGHDTNLANLGGALELAWNTLPGQDNTPPGGELVFERWRRLSDNSQWIOVSLVFTQLQOM 360
 QY 361 RDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 410
 Db 361 RDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 410

RESULT 2
 US-10-021-723A-15
 ; Sequence 15, Application US/10021723A
 ; Publication No. US20030101476A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Short, Jay
 ; APPLICANT: Mathur, Eric
 ; APPLICANT: Richardson, Toby
 ; APPLICANT: Robertson, Dan
 ; APPLICANT: Barton, Nelson
 ; TITLE OF INVENTION: Recombinant Phytases and Uses Thereof
 ; FILE REFERENCE: 112766.140 (DIV-016CIP)
 ; CURRENT APPLICATION NUMBER: US/10/021,723A
 ; CURRENT FILING DATE: 2002-10-23
 ; PRIOR APPLICATION NUMBER: US 60/255,090
 ; PRIOR FILING DATE: 2000-12-12
 ; NUMBER OF SEQ ID NOS: 74
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 15
 ; LENGTH: 410
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: amino acid sequence of a phytase enzyme with
 ; OTHER INFORMATION: predicted glycosylation sites
 US-10-021-723A-15

Query Match 100.0%; Score 2153; DB 14; Length 410;
 Best Local Similarity 100.0%; Pred. No. 2.3e-199;
 Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 QSEPELKLESVIVSRHGVRAPTKATQLMQDVTPDAMPTWPKLGELTPRGGLIAYLGH 60
 Db 1 QSEPELKLESVIVSRHGVRAPTKATQLMQDVTPDAMPTWPKLGELTPRGGLIAYLGH 60
 QY 61 YWRQLVADGLLPKCGCPQSGQVAILADVDERTRKTGEAFAGLAPDCAITVHTQADTSS 120
 Db 61 YWRQLVADGLLPKCGCPQSGQVAILADVDERTRKTGEAFAGLAPDCAITVHTQADTSS 120
 QY 121 PDPLFNPLKTVGCOLDNANVTDAILERAGGSIADFTGHYQTAFARELERVLPQSNLCLK 180
 Db 121 PDPLFNPLKTVGCOLDNANVTDAILERAGGSIADFTGHYQTAFARELERVLPQSNLCLK 180
 QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 240
 Db 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 240
 QY 241 HOWNTLLSLHNAQFDLLQRTPEVARSRATPLDLIKTALTALTPHPPOKQAYGVTLPSTVLF 300
 Db 241 HOWNTLLSLHNAQFDLLQRTPEVARSRATPLDLIKTALTALTPHPPOKQAYGVTLPSTVLF 300

QY 301 AGHDTNLANLGGALELAWNTLPGQDNTPPGGELVFERWRRLSDNSQWIOVSLVFTQLQOM 360
 Db 301 AGHDTNLANLGGALELAWNTLPGQDNTPPGGELVFERWRRLSDNSQWIOVSLVFTQLQOM 360
 QY 361 RDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 410
 Db 361 RDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 410

RESULT 3
 US-10-334-672-1
 ; Sequence 1, Application US/10334672
 ; Publication No. US20030157646A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lanahan, Mike
 ; APPLICANT: Koepf, Edward
 ; APPLICANT: Kretz, Keith
 ; TITLE OF INVENTION: Microbially-Expressed Thermotolerant Phytase For Animal Feed
 ; FILE REFERENCE: SYNG-P01-001
 ; CURRENT APPLICATION NUMBER: US/10/334,672
 ; CURRENT FILING DATE: 2002-12-30
 ; PRIOR APPLICATION NUMBER: 60/344,523
 ; PRIOR FILING DATE: 2001-12-28
 ; NUMBER OF SEQ ID NOS: 4
 ; SOFTWARE: Patentin version 3.1
 ; SEQ ID NO 1
 ; LENGTH: 412
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: No. US20030157646A19X Phytase
 US-10-334-672-1

Query Match 100.0%; Score 2153; DB 14; Length 412;
 Best Local Similarity 100.0%; Pred. No. 2.3e-199;
 Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 QSEPELKLESVIVSRHGVRAPTKATQLMQDVTPDAMPTWPKLGELTPRGGLIAYLGH 60
 Db 3 QSEPELKLESVIVSRHGVRAPTKATQLMQDVTPDAMPTWPKLGELTPRGGLIAYLGH 62
 QY 61 YWRQLVADGLLPKCGCPQSGQVAILADVDERTRKTGEAFAGLAPDCAITVHTQADTSS 120
 Db 63 YWRQLVADGLLPKCGCPQSGQVAILADVDERTRKTGEAFAGLAPDCAITVHTQADTSS 122
 QY 121 PDPLFNPLKTVGCOLDNANVTDAILERAGGSIADFTGHYQTAFARELERVLPQSNLCLK 180
 Db 123 PDPLFNPLKTVGCOLDNANVTDAILERAGGSIADFTGHYQTAFARELERVLPQSNLCLK 182
 QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 240
 Db 183 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 242
 QY 241 HOWNTLLSLHNAQFDLLQRTPEVARSRATPLDLIKTALTALTPHPPOKQAYGVTLPSTVLF 300
 Db 243 HOWNTLLSLHNAQFDLLQRTPEVARSRATPLDLIKTALTALTPHPPOKQAYGVTLPSTVLF 302
 QY 301 AGHDTNLANLGGALELAWNTLPGQDNTPPGGELVFERWRRLSDNSQWIOVSLVFTQLQOM 360
 Db 303 AGHDTNLANLGGALELAWNTLPGQDNTPPGGELVFERWRRLSDNSQWIOVSLVFTQLQOM 362
 QY 361 RDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 410
 Db 363 RDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 412

RESULT 4
 US-10-334-671-1
 ; Sequence 1, Application US/10334671
 ; Publication No. US20030170293A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lanahan, Mike
 ; APPLICANT: Koepf, Edward

APPLICANT: Kretz, Keith
TITLE OF INVENTION: Microbially-Expressed Thermotolerant Phytase For Animal Feed
FILE REFERENCE: 70098
CURRENT APPLICATION NUMBER: US/10/334,671
CURRENT FILING DATE: 2002-12-30
PRIOR APPLICATION NUMBER: 60/344,523
PRIOR FILING DATE: 2001-12-28
NUMBER OF SEQ ID NOS: 4
SOFTWARE: Patent in version 3.1
SEQ ID NO 1
LENGTH: 412
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: No. US20030170293A19X Phytase
US-10-334-671-1

Query Match 100.0%; Score 2153; DB 14; Length 412;
Best Local Similarity 100.0%; Pred. No. 2.3e-199;
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QSEPELKLESVIVSRHGVRAPTKATQLMQDVTDPDAMPTWPKLGELTPRGGLIAYLGH 60
DB 3 QSEPELKLESVIVSRHGVRAPTKATQLMQDVTDPDAMPTWPKLGELTPRGGLIAYLGH 62
QY 61 YWRQLVADGLLPKCGCPQSGQVIAIADVDERTRKTGEAFAAGLAPDCAITVHTQADTSS 120
DB 63 YWRQLVADGLLPKCGCPQSGQVIAIADVDERTRKTGEAFAAGLAPDCAITVHTQADTSS 122
QY 121 PDPLFNPLKTVGCOLDNANVTDAILERAGGSIAFTGHYQTAFRELERVLPQSNLCLK 180
DB 123 PDPLFNPLKTVGCOLDNANVTDAILERAGGSIAFTGHYQTAFRELERVLPQSNLCLK 182
QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLQQAQGMPEPGWGRITDS 240
DB 183 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLQQAQGMPEPGWGRITDS 242
QY 241 HQWNTLLSHNAQFDLLQRTPEVARSRATPLDLIKTALTALTPHPKQKQAYGVTLPTSVLFI 300
DB 243 HQWNTLLSHNAQFDLLQRTPEVARSRATPLDLIKTALTALTPHPKQKQAYGVTLPTSVLFI 302
QY 301 AGHDTNLNLGGALELNWTLPGQPDNTPPGGELVFERWRRLSDNSQIQLSVFQTLQOM 360
DB 303 AGHDTNLNLGGALELNWTLPGQPDNTPPGGELVFERWRRLSDNSQIQLSVFQTLQOM 362
QY 361 RDKTPLSLNTPPEGVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 410
DB 363 RDKTPLSLNTPPEGVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 412

RESULT 5
US-10-156-660-2
Sequence 2, Application US/10156660
Publication No. US20030103958A1
GENERAL INFORMATION:
APPLICANT: Short, Jay M.
APPLICANT: Kretz, Keith
APPLICANT: Gray, Kevin A.
APPLICANT: Barton, Nelson R.
APPLICANT: Garrett, James B.
APPLICANT: O'Donoghue, Eileen
APPLICANT: Mathur, Eric J.
TITLE OF INVENTION: PHYTASES, NUCLEIC ACIDS ENCODING THEM
TITLE OF INVENTION: AND METHODS FOR MAKING AND USING THEM
FILE REFERENCE: 09010-029007
CURRENT APPLICATION NUMBER: US/10/156,660
CURRENT FILING DATE: 2002-10-01
PRIOR APPLICATION NUMBER: US 09/866,379
PRIOR FILING DATE: 2001-05-24
PRIOR APPLICATION NUMBER: US 09/580,515
PRIOR FILING DATE: 2000-05-25
PRIOR APPLICATION NUMBER: US 09/318,528
PRIOR FILING DATE: 1999-05-25

PRIOR APPLICATION NUMBER: US 09/291,931
PRIOR FILING DATE: 1999-04-13
PRIOR APPLICATION NUMBER: US 09/259,214
PRIOR FILING DATE: 1999-03-01
PRIOR APPLICATION NUMBER: US 08/910,798
PRIOR FILING DATE: 1997-08-13
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 436
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: modified phytase enzyme
US-10-156-660-2

Query Match 100.0%; Score 2153; DB 14; Length 436;
Best Local Similarity 100.0%; Pred. No. 2.6e-199;
Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QSEPELKLESVIVSRHGVRAPTKATQLMQDVTDPDAMPTWPKLGELTPRGGLIAYLGH 60
DB 23 QSEPELKLESVIVSRHGVRAPTKATQLMQDVTDPDAMPTWPKLGELTPRGGLIAYLGH 82
QY 61 YWRQLVADGLLPKCGCPQSGQVIAIADVDERTRKTGEAFAAGLAPDCAITVHTQADTSS 120
DB 83 YWRQLVADGLLPKCGCPQSGQVIAIADVDERTRKTGEAFAAGLAPDCAITVHTQADTSS 142
QY 121 PDPLFNPLKTVGCOLDNANVTDAILERAGGSIAFTGHYQTAFRELERVLPQSNLCLK 180
DB 143 PDPLFNPLKTVGCOLDNANVTDAILERAGGSIAFTGHYQTAFRELERVLPQSNLCLK 202
QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLQQAQGMPEPGWGRITDS 240
DB 203 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLQQAQGMPEPGWGRITDS 262
QY 241 HQWNTLLSHNAQFDLLQRTPEVARSRATPLDLIKTALTALTPHPKQKQAYGVTLPTSVLFI 300
DB 263 HQWNTLLSHNAQFDLLQRTPEVARSRATPLDLIKTALTALTPHPKQKQAYGVTLPTSVLFI 322
QY 301 AGHDTNLNLGGALELNWTLPGQPDNTPPGGELVFERWRRLSDNSQIQLSVFQTLQOM 360
DB 323 AGHDTNLNLGGALELNWTLPGQPDNTPPGGELVFERWRRLSDNSQIQLSVFQTLQOM 382
QY 361 RDKTPLSLNTPPEGVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 410
DB 383 RDKTPLSLNTPPEGVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 432

RESULT 6
US-09-866-379-10
Sequence 10, Application US/09866379
Patent No. US20020136754A1
GENERAL INFORMATION:
APPLICANT: DIVERSA CORPORATION
APPLICANT: SHORT, Jay
APPLICANT: KRETZ, Keith
APPLICANT: GRAY, Kevin
APPLICANT: BARTON, Nelson
APPLICANT: GARRETT, James
APPLICANT: O'DONOGHUE, Eileen
TITLE OF INVENTION: RECOMBINANT BACTERIAL PHYTASES AND USES THEREOF
FILE REFERENCE: DIVER1370-7
CURRENT APPLICATION NUMBER: US/09/866,379
CURRENT FILING DATE: 2001-05-24
PRIOR APPLICATION NUMBER: US 09/580,515
PRIOR FILING DATE: 2000-05-25
PRIOR APPLICATION NUMBER: US 09/318,528
PRIOR FILING DATE: 1999-05-25
PRIOR APPLICATION NUMBER: US 09/291,931
PRIOR FILING DATE: 1999-04-13
PRIOR APPLICATION NUMBER: US 09/259,214
PRIOR FILING DATE: 1999-03-01

; PRIOR APPLICATION NUMBER: US 08/910,798
; PRIOR FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 10
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Modified phytase
US-09-866-379-10

Query Match 98.6%; Score 2123; DB 9; Length 430;
Best Local Similarity 99.5%; Pred. No. 2e-196;
Matches 408; Conservative 0; Mismatches 0; Indels 2; Gaps 2;
QY 1 QSEPELKLESVVIYSRHGVRAPTKATQMQDVTDPDAMPTWPKLGELTPRGGLIAYLGH 60
DB 23 QSEPELKLESVVIYSRHGVRAPTKATQMQDVTDPDAMPTWPKLGELTPRGGLIAYLGH 82
QY 61 YWRQLVADGLLPKCGCPQSGQVAIIADVDERTKRTGTEFAAGLAPDCAITVHTQADTSS 120
DB 83 YWRQLVADGLLPKCGCPQSGQVAIIADVDERTKRTGTEFAAGLAPDCAITVHTQADTSS 142
QY 121 PDPLFNPLKTGVCOLDNANVTDAILLERAGGSADFTGHYQTAFARELERVLPQSNLCLK 180
DB 143 PDPLFNPLKTGVCOLDNANVTDAILE-AGGSADFTGHYQTAFARELERVLPQSNLCLK 201
QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 240
DB 202 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 261
QY 241 HQWNTLLSLHNAQFDLQRTPEVARSRATPLLDLIKTALTTPHPPOKQAYGVTLPSTVLFI 300
DB 262 HQWNTLLSLHNAQFD-LQRTPEVARSRATPLLDLIKTALTTPHPPOKQAYGVTLPSTVLFI 320
QY 301 AGHDNTNLNLGALLENLNTLPGQPDNTPPGGELVFERWRRLSDNSQWIQVSLVFQTLQOM 360
DB 321 AGHDNTNLNLGALLENLNTLPGQPDNTPPGGELVFERWRRLSDNSQWIQVSLVFQTLQOM 380
QY 361 RDKTPLSLNTPPGVKLTLAGCEERNAQCMCSLAGFTQIVNEARIPACSL 410
DB 381 RDKTPLSLNTPPGVKLTLAGCEERNAQCMCSLAGFTQIVNEARIPACSL 430

RESULT 7
US-09-866-379-8
; Sequence 8, Application US/09866379
; Patent No. US20020136754A1
; GENERAL INFORMATION:
; APPLICANT: DIVERSA CORPORATION
; APPLICANT: SHORT, Jay
; APPLICANT: KRETZ, Keith
; APPLICANT: GRAY, Kevin
; APPLICANT: BARTON, Nelson
; APPLICANT: GARRETT, James
; APPLICANT: O'DONOGHUE, Eileen
; TITLE OF INVENTION: RECOMBINANT BACTERIAL PHYTASES AND USES THEREOF
; FILE REFERENCE: DIVER1370-7
; CURRENT APPLICATION NUMBER: US 09/866,379
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 09/580,515
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 09/318,528
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: US 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: US 09/259,214
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: US 08/910,798
; PRIOR FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent in version 3.1

; SEQ ID NO 8
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-866-379-8
Query Match 96.5%; Score 2077; DB 9; Length 432;
Best Local Similarity 98.0%; Pred. No. 5.7e-192;
Matches 402; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 1 QSEPELKLESVVIYSRHGVRAPTKATQMQDVTDPDAMPTWPKLGELTPRGGLIAYLGH 60
DB 23 QSEPELKLESVVIYSRHGVRAPTKATQMQDVTDPDAMPTWPKLGELTPRGGLIAYLGH 82
QY 61 YWRQLVADGLLPKCGCPQSGQVAIIADVDERTKRTGTEFAAGLAPDCAITVHTQADTSS 120
DB 83 YWRQLVADGLLPKCGCPQSGQVAIIADVDERTKRTGTEFAAGLAPDCAITVHTQADTSS 142
QY 121 PDPLFNPLKTGVCOLDNANVTDAILLERAGGSADFTGHYQTAFARELERVLPQSNLCLK 180
DB 143 PDPLFNPLKTGVCOLDNANVTDAILLERAGGSADFTGHYQTAFARELERVLPQSNLCLK 202
QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 240
DB 203 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 262
QY 241 HQWNTLLSLHNAQFDLQRTPEVARSRATPLLDLIKTALTTPHPPOKQAYGVTLPSTVLFI 300
DB 263 HQWNTLLSLHNAQFDLQRTPEVARSRATPLLDLIKTALTTPHPPOKQAYGVTLPSTVLFI 322
QY 301 AGHDNTNLNLGALLENLNTLPGQPDNTPPGGELVFERWRRLSDNSQWIQVSLVFQTLQOM 360
DB 323 AGHDNTNLNLGALLENLNTLPGQPDNTPPGGELVFERWRRLSDNSQWIQVSLVFQTLQOM 382
QY 361 RDKTPLSLNTPPGVKLTLAGCEERNAQCMCSLAGFTQIVNEARIPACSL 410
DB 383 RDKTPLSLNTPPGVKLTLAGCEERNAQCMCSLAGFTQIVNEARIPACSL 432

RESULT 8
US-10-282-122A-43351
; Sequence 43351, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Onisen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625


```

; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 43351
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-282-122A-43351

Query Match      96.5%; Score 2077; DB 12; Length 432;
Best Local Similarity 98.0%; Pred. No. 5.7e-192;
Matches 402; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 QSEPELKLESVIVSRHGVRAPTKATQMQDVTDPDWTWPKLGELTPRGGLIAYLGH 60
DB 23 QSEPELKLESVIVSRHGVRAPTKATQMQDVTDPDWTWPKLGELTPRGGLIAYLGH 82
QY 61 YRQRLVADGLLPKCCPQSGQVAIIADVDERTRKTGEAFAAGLAPDCAITVHTQADTSS 120
DB 83 YRQRLVADGLLPKCCPQSGQVAIIADVDERTRKTGEAFAAGLAPDCAITVHTQADTSS 142
QY 121 PDPLENPLKTGVCQOLDNANVTDAIILRAGGSADFTGHYQTAFRELERVLNFPQSNCLK 180
DB 143 PDPLENPLKTGVCQOLDNANVTDAIILRAGGSADFTGHYQTAFRELERVLNFPQSNCLK 202
QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMPGPGWGRITDS 240
DB 203 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMPGPGWGRITDS 262
QY 241 HOWNILLSLHNAQFILLQRTPEVARSRATPLDLIKTALTTPHPPOKQAYGVTLPSTVLF 300
DB 263 HOWNILLSLHNAQFILLQRTPEVARSRATPLDLIKTALTTPHPPOKQAYGVTLPSTVLF 322
QY 301 AGHDTNLNLGALNLTLPQPDNTPPGGELVFERWRRLSDNSQWQVSLVFTQLQOM 360
DB 323 AGHDTNLNLGALNLTLPQPDNTPPGGELVFERWRRLSDNSQWQVSLVFTQLQOM 382
QY 361 RDKTPLSLNTPPGVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 410
DB 383 RDKTPLSLNTPPGVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 432

RESULT 9
US-10-156-660-4
; Sequence 4, Application US/10156660
; Publication No. US20030103958A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay M.
; APPLICANT: Kretz, Kevin A.
; APPLICANT: Gray, Kevin A.
; APPLICANT: Barton, Nelson R.
; APPLICANT: Garrett, James B.
; APPLICANT: O'Donoghue, Eileen
; APPLICANT: Mathur, Eric J.
; TITLE OF INVENTION: PHYTASES, NUCLEIC ACIDS ENCODING THEM
; TITLE OF INVENTION: AND METHODS FOR MAKING AND USING THEM
; FILE REFERENCE: 09010-029007
; CURRENT APPLICATION NUMBER: US/10156,660
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: US 09/866,379
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 09/580,515
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 09/318,528
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: US 09/291,931

```

```

; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: US 09/259,214
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: US 08/910,798
; PRIOR FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-156-660-4

Query Match      96.5%; Score 2077; DB 14; Length 432;
Best Local Similarity 98.0%; Pred. No. 5.7e-192;
Matches 402; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 QSEPELKLESVIVSRHGVRAPTKATQMQDVTDPDWTWPKLGELTPRGGLIAYLGH 60
DB 23 QSEPELKLESVIVSRHGVRAPTKATQMQDVTDPDWTWPKLGELTPRGGLIAYLGH 82
QY 61 YRQRLVADGLLPKCCPQSGQVAIIADVDERTRKTGEAFAAGLAPDCAITVHTQADTSS 120
DB 83 YRQRLVADGLLPKCCPQSGQVAIIADVDERTRKTGEAFAAGLAPDCAITVHTQADTSS 142
QY 121 PDPLENPLKTGVCQOLDNANVTDAIILRAGGSADFTGHYQTAFRELERVLNFPQSNCLK 180
DB 143 PDPLENPLKTGVCQOLDNANVTDAIILRAGGSADFTGHYQTAFRELERVLNFPQSNCLK 202
QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMPGPGWGRITDS 240
DB 203 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMPGPGWGRITDS 262
QY 241 HOWNILLSLHNAQFILLQRTPEVARSRATPLDLIKTALTTPHPPOKQAYGVTLPSTVLF 300
DB 263 HOWNILLSLHNAQFILLQRTPEVARSRATPLDLIKTALTTPHPPOKQAYGVTLPSTVLF 322
QY 301 AGHDTNLNLGALNLTLPQPDNTPPGGELVFERWRRLSDNSQWQVSLVFTQLQOM 360
DB 323 AGHDTNLNLGALNLTLPQPDNTPPGGELVFERWRRLSDNSQWQVSLVFTQLQOM 382
QY 361 RDKTPLSLNTPPGVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 410
DB 383 RDKTPLSLNTPPGVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 432

RESULT 10
US-09-777-566A-2
; Sequence 2, Application US/09777566A
; Patent No. US20010055788A1
; GENERAL INFORMATION:
; APPLICANT: DIVERSA CORPORATION
; APPLICANT: Short, Jay
; APPLICANT: Kretz, Keith
; TITLE OF INVENTION: RECOMBINANT BACTERIAL PHYTASES AND USES THEREOF
; FILE REFERENCE: DIVER1370-6
; CURRENT APPLICATION NUMBER: US/09/777,566A
; CURRENT FILING DATE: 2001-06-11
; PRIOR APPLICATION NUMBER: US 09/318,528
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: US 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: US 09/259,214
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: US 08/910,798
; PRIOR FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:

```

NAME/KEY: misc.feature
LOCATION: (1)..(1323)
OTHER INFORMATION: n is any nucleotide
US-09-777-566A-2

Query Match 96.0%; Score 2066; DB 9; Length 440;
Best Local Similarity 97.6%; Pred. No. 6.8e-191;
Matches 400; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
QY 1 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTDPDAMPTWPKLGTGELTPRGGELIAYLGH 60
DB 23 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTDPDAMPTWPKLGTGELTPRGGELIAYLGH 82
QY 61 YRQRLVADGLLPKCCGQSQGVALLIADVDERTRKTGEAFAGLAPDCAITVHTQADTSS 120
DB 83 YRQRLVADGLLAKKGCQSQGVALLIADVDERTRKTGEAFAGLAPDCAITVHTQADTSS 142
QY 121 PDPLFNPLKTGVCQLDNANVTDAILERAGGSIADFTGHYQTAFRELERVLPQSNLCLK 180
DB 143 PDPLFNPLKTGVCQLDNANVTDAILERAGGSIADFTGHYQTAFRELERVLPQSNLCLK 202
QY 181 REKQDESSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 240
DB 203 REKQDESSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 262
QY 241 HQWNTLLSHNAQFVLLQRTPEVARSRAATPLLDLTKALTTPHPPQKQAYGVTLPTSVLFI 300
DB 263 HQWNTLLSHNAQFVLLQRTPEVARSRAATPLLDLTKALTTPHPPQKQAYGVTLPTSVLFI 322
QY 301 AGHDTNLNLGALBELNWTLPQPDNTPPGGELVFERWRRLSDNSQIQLSVLFTQLOQM 360
DB 323 AGHDTNLNLGALBELNWTLPQPDNTPPGGELVFERWRRLSDNSQIQLSVLFTQLOQM 382
QY 361 RDKTPLSLNTPPGEVKLTLAGCEERNAQCMCSLAGFTQIVNEARIPACSL 410
DB 383 RDKTPLSLNTPPGEVKLTLAGCEERNAQCMCSLAGFTQIVNEARIPACSL 432

RESULT 11
US-09-866-379-2
Sequence 2, Application US/09866379
Patent No. US20020136754A1
GENERAL INFORMATION:
APPLICANT: DIVERSA CORPORATION
APPLICANT: SHORT, Jay
APPLICANT: KRETZ, Keith
APPLICANT: GRAY, Kevin
APPLICANT: BARTON, Nelson
APPLICANT: GARRETT, James
APPLICANT: O'DONOGHUE, Eileen
TITLE OF INVENTION: RECOMBINANT BACTERIAL PHYTASES AND USES THEREOF
FILE REFERENCE: DIVER1370-7
CURRENT APPLICATION NUMBER: US/09/866,379
PRIOR FILING DATE: 2001-05-24
PRIOR APPLICATION NUMBER: US 09/580,515
PRIOR FILING DATE: 2000-05-25
PRIOR APPLICATION NUMBER: US 09/318,528
PRIOR FILING DATE: 1999-05-25
PRIOR APPLICATION NUMBER: US 09/291,931
PRIOR FILING DATE: 1999-04-13
PRIOR APPLICATION NUMBER: US 09/259,214
PRIOR FILING DATE: 1999-03-01
PRIOR APPLICATION NUMBER: US 08/910,798
PRIOR FILING DATE: 1997-08-13
NUMBER OF SEQ ID NOS: 10
SOFTWARE: Patent in version 3.1
SEQ ID NO 2
LENGTH: 440
TYPE: PRT
ORGANISM: Escherichia coli
US-09-866-379-2

Query Match 96.0%; Score 2066; DB 9; Length 440;

Best Local Similarity 97.6%; Pred. No. 6.8e-191;
Matches 400; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
QY 1 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTDPDAMPTWPKLGTGELTPRGGELIAYLGH 60
DB 23 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTDPDAMPTWPKLGTGELTPRGGELIAYLGH 82
QY 61 YRQRLVADGLLPKCCGQSQGVALLIADVDERTRKTGEAFAGLAPDCAITVHTQADTSS 120
DB 83 YRQRLVADGLLAKKGCQSQGVALLIADVDERTRKTGEAFAGLAPDCAITVHTQADTSS 142
QY 121 PDPLFNPLKTGVCQLDNANVTDAILERAGGSIADFTGHYQTAFRELERVLPQSNLCLK 180
DB 143 PDPLFNPLKTGVCQLDNANVTDAILERAGGSIADFTGHYQTAFRELERVLPQSNLCLK 202
QY 181 REKQDESSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 240
DB 203 REKQDESSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 262
QY 241 HQWNTLLSHNAQFVLLQRTPEVARSRAATPLLDLTKALTTPHPPQKQAYGVTLPTSVLFI 300
DB 263 HQWNTLLSHNAQFVLLQRTPEVARSRAATPLLDLTKALTTPHPPQKQAYGVTLPTSVLFI 322
QY 301 AGHDTNLNLGALBELNWTLPQPDNTPPGGELVFERWRRLSDNSQIQLSVLFTQLOQM 360
DB 323 AGHDTNLNLGALBELNWTLPQPDNTPPGGELVFERWRRLSDNSQIQLSVLFTQLOQM 382
QY 361 RDKTPLSLNTPPGEVKLTLAGCEERNAQCMCSLAGFTQIVNEARIPACSL 410
DB 383 RDKTPLSLNTPPGEVKLTLAGCEERNAQCMCSLAGFTQIVNEARIPACSL 432

RESULT 12
US-10-034-985-2
Sequence 2, Application US/10034985
Publication No. US20030049815A1
GENERAL INFORMATION:
APPLICANT: Kretz, Keith
TITLE OF INVENTION: NOVEL PHYTASE
FILE REFERENCE: 09010/029003
CURRENT APPLICATION NUMBER: US/10/034,985
CURRENT FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: US/09/580,515
PRIOR FILING DATE: 1999-05-25
PRIOR APPLICATION NUMBER: 09/291,931
PRIOR FILING DATE: 1999-04-13
PRIOR APPLICATION NUMBER: 08/910,798
PRIOR FILING DATE: 1997-08-13
PRIOR APPLICATION NUMBER: 09/259,214
PRIOR FILING DATE: 1999-03-01
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 440
TYPE: PRT
ORGANISM: Escherichia coli
US-10-034-985-2

Query Match 96.0%; Score 2066; DB 14; Length 440;
Best Local Similarity 97.6%; Pred. No. 6.8e-191;
Matches 400; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
QY 1 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTDPDAMPTWPKLGTGELTPRGGELIAYLGH 60
DB 23 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTDPDAMPTWPKLGTGELTPRGGELIAYLGH 82
QY 61 YRQRLVADGLLPKCCGQSQGVALLIADVDERTRKTGEAFAGLAPDCAITVHTQADTSS 120
DB 83 YRQRLVADGLLAKKGCQSQGVALLIADVDERTRKTGEAFAGLAPDCAITVHTQADTSS 142
QY 121 PDPLFNPLKTGVCQLDNANVTDAILERAGGSIADFTGHYQTAFRELERVLPQSNLCLK 180
DB 143 PDPLFNPLKTGVCQLDNANVTDAILERAGGSIADFTGHYQTAFRELERVLPQSNLCLK 202

```
QY 181 REKQDECSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMPFGWGRITDS 240
DB 203 REKQDECSLTQALPSELKVSADNVSLTGAVSLASMLTEIFLLQQAQMPFGWGRITDS 262
QY 241 HQWNTLLSLHNAQFLLQRTPEVARSRAATPLDLIKTALTTPHPKQKQAYGVTLPTSVLFI 300
DB 263 HQWNTLLSLHNAQFLLQRTPEVARSRAATPLDLIIMALTTPHPKQKQAYGVTLPTSVLFI 322
QY 301 AGHDTNLANLGGALELNWTLPGQPDNTPPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 360
DB 323 AGHDTNLANLGGALELNWTLPGQPDNTPPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 382
QY 361 RDKTPLSLNTPPGEVKLTLAGCEERNAQCMCSLAGFTQIVNEARIPACSL 410
DB 383 RDKTPLSLNTPPGEVKLTLAGCEERNAQCMCSLAGFTQIVNEARIPACSL 432

RESULT 13
US-10-430-356-2
; Sequence 2, Application US/10430356
; Publication No. US20030232041A1
; GENERAL INFORMATION:
; APPLICANT: Kretz, Keith
; TITLE OF INVENTION: NOVEL PHYTASE
; FILE REFERENCE: 09010/029003
; CURRENT APPLICATION NUMBER: US/10/430,356
; CURRENT FILING DATE: 2003-05-05
; PRIOR APPLICATION NUMBER: US/09/580,515
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: 09/318,528
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: 08/910,798
; PRIOR FILING DATE: 1997-08-13
; PRIOR APPLICATION NUMBER: 09/259,214
; PRIOR FILING DATE: 1999-03-01
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-430-356-2

Query Match 96.0%; Score 2066; DB 15; Length 440;
Best Local Similarity 97.6%; Pred. No. 6.8e-191;
Matches 400; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 OSEPELKLESVVIVSRHGVRAPTATQALMDQVTPDAWPTWPKLGELTPRGGLIAYLGH 60
DB 23 OSEPELKLESVVIVSRHGVRAPTATQALMDQVTPDAWPTWPKLGELTPRGGLIAYLGH 82
QY 61 YWRQLVADGLLPKCGCPQSGQVAILADVDERTRKTGEAFAAGLAPDCAITVHTQADTSS 120
DB 83 YWRQLVADGLLAKKCGCPQSGQVAILADVDERTRKTGEAFAAGLAPDCAITVHTQADTSS 142
QY 121 PDPLFNPLKTGVCQLDANVTDAIILRAGGSIADFTGHYQTAFRELERVLPFQSNCLK 180
DB 143 PDPLFNPLKTGVCQLDANVTDAIILRAGGSIADFTGHYQTAFRELERVLPFQSNCLK 202
QY 181 REKQDECSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMPFGWGRITDS 240
DB 203 REKQDECSLTQALPSELKVSADNVSLTGAVSLASMLTEIFLLQQAQMPFGWGRITDS 262
QY 241 HQWNTLLSLHNAQFLLQRTPEVARSRAATPLDLIKTALTTPHPKQKQAYGVTLPTSVLFI 300
DB 263 HQWNTLLSLHNAQFLLQRTPEVARSRAATPLDLIIMALTTPHPKQKQAYGVTLPTSVLFI 322
QY 301 AGHDTNLANLGGALELNWTLPGQPDNTPPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 360
DB 323 AGHDTNLANLGGALELNWTLPGQPDNTPPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 382

RESULT 14
US-10-284-962-3
; Sequence 3, Application US/10284962
; Publication No. US20030206913A1
; GENERAL INFORMATION:
; APPLICANT: Weibel, Douglas M.
; APPLICANT: Orr, Donald E.
; APPLICANT: Ruch, Frank E.
; TITLE OF INVENTION: PHYTASE-CONTAINING ANIMAL FOOD AND METHOD
; FILE REFERENCE: 834460-71725
; CURRENT APPLICATION NUMBER: US/10/284,962
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: US 60/335,303
; PRIOR FILING DATE: 2001-10-31
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-284-962-3

Query Match 94.8%; Score 2042; DB 15; Length 432;
Best Local Similarity 96.6%; Pred. No. 1.4e-188;
Matches 396; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 QSEPELKLESVVIVSRHGVRAPTATQALMDQVTPDAWPTWPKLGELTPRGGLIAYLGH 60
DB 23 QSEPELKLESVVIVSRHGVRAPTATQALMDQVTPDAWPTWPKLGELTPRGGLIAYLGH 82
QY 61 YWRQLVADGLLPKCGCPQSGQVAILADVDERTRKTGEAFAAGLAPDCAITVHTQADTSS 120
DB 83 YWRQLVADGLLAKKCGCPQSGQVAILADVDERTRKTGEAFAAGLAPDCAITVHTQADTSS 142
QY 121 PDPLFNPLKTGVCQLDANVTDAIILRAGGSIADFTGHYQTAFRELERVLPFQSNCLK 180
DB 143 PDPLFNPLKTGVCQLDANVTDAIILRAGGSIADFTGHYQTAFRELERVLPFQSNCLK 202
QY 181 REKQDECSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMPFGWGRITDS 240
DB 203 REKQDECSLTQALPSELKVSADNVSLTGAVSLASMLTEIFLLQQAQMPFGWGRITDS 262
QY 241 HQWNTLLSLHNAQFLLQRTPEVARSRAATPLDLIKTALTTPHPKQKQAYGVTLPTSVLFI 300
DB 263 HQWNTLLSLHNAQFLLQRTPEVARSRAATPLDLIIMALTTPHPKQKQAYGVTLPTSVLFI 322
QY 301 AGHDTNLANLGGALELNWTLPGQPDNTPPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 360
DB 323 AGHDTNLANLGGALELNWTLPGQPDNTPPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 382

RESULT 15
US-10-284-962-14
; Sequence 14, Application US/10284962
; Publication No. US20030206913A1
; GENERAL INFORMATION:
; APPLICANT: Weibel, Douglas M.
; APPLICANT: Orr, Donald E.
; APPLICANT: Ruch, Frank E.
; TITLE OF INVENTION: PHYTASE-CONTAINING ANIMAL FOOD AND METHOD
; FILE REFERENCE: 834460-71725
; CURRENT APPLICATION NUMBER: US/10/284,962
; CURRENT FILING DATE: 2002-10-31
```

```
QY 361 RDKTPLSLNTPPGEVKLTLAGCEERNAQCMCSLAGFTQIVNEARIPACSL 410
DB 383 RDKTPLSLNTPPGEVKLTLAGCEERNAQCMCSLAGFTQIVNEARIPACSL 432

RESULT 14
US-10-284-962-3
; Sequence 3, Application US/10284962
; Publication No. US20030206913A1
; GENERAL INFORMATION:
; APPLICANT: Weibel, Douglas M.
; APPLICANT: Orr, Donald E.
; APPLICANT: Ruch, Frank E.
; TITLE OF INVENTION: PHYTASE-CONTAINING ANIMAL FOOD AND METHOD
; FILE REFERENCE: 834460-71725
; CURRENT APPLICATION NUMBER: US/10/284,962
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: US 60/335,303
; PRIOR FILING DATE: 2001-10-31
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-284-962-3

Query Match 94.8%; Score 2042; DB 15; Length 432;
Best Local Similarity 96.6%; Pred. No. 1.4e-188;
Matches 396; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 QSEPELKLESVVIVSRHGVRAPTATQALMDQVTPDAWPTWPKLGELTPRGGLIAYLGH 60
DB 23 QSEPELKLESVVIVSRHGVRAPTATQALMDQVTPDAWPTWPKLGELTPRGGLIAYLGH 82
QY 61 YWRQLVADGLLPKCGCPQSGQVAILADVDERTRKTGEAFAAGLAPDCAITVHTQADTSS 120
DB 83 YWRQLVADGLLAKKCGCPQSGQVAILADVDERTRKTGEAFAAGLAPDCAITVHTQADTSS 142
QY 121 PDPLFNPLKTGVCQLDANVTDAIILRAGGSIADFTGHYQTAFRELERVLPFQSNCLK 180
DB 143 PDPLFNPLKTGVCQLDANVTDAIILRAGGSIADFTGHYQTAFRELERVLPFQSNCLK 202
QY 181 REKQDECSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMPFGWGRITDS 240
DB 203 REKQDECSLTQALPSELKVSADNVSLTGAVSLASMLTEIFLLQQAQMPFGWGRITDS 262
QY 241 HQWNTLLSLHNAQFLLQRTPEVARSRAATPLDLIKTALTTPHPKQKQAYGVTLPTSVLFI 300
DB 263 HQWNTLLSLHNAQFLLQRTPEVARSRAATPLDLIIMALTTPHPKQKQAYGVTLPTSVLFI 322
QY 301 AGHDTNLANLGGALELNWTLPGQPDNTPPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 360
DB 323 AGHDTNLANLGGALELNWTLPGQPDNTPPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 382

RESULT 15
US-10-284-962-14
; Sequence 14, Application US/10284962
; Publication No. US20030206913A1
; GENERAL INFORMATION:
; APPLICANT: Weibel, Douglas M.
; APPLICANT: Orr, Donald E.
; APPLICANT: Ruch, Frank E.
; TITLE OF INVENTION: PHYTASE-CONTAINING ANIMAL FOOD AND METHOD
; FILE REFERENCE: 834460-71725
; CURRENT APPLICATION NUMBER: US/10/284,962
; CURRENT FILING DATE: 2002-10-31
```

```
; PRIOR APPLICATION NUMBER: US 60/335,303
; PRIOR FILING DATE: 2001-10-31
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-284-962-14

Query Match      94.8%; Score 2042; DB 15; Length 432;
Best Local Similarity 96.6%; Pred. No. 1.4e-188;
Matches 396; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 QSEPEIKLESVVIVSRHGVRAPTKATQIMQDVTPDAMPTWPKIGELTPRGELIAYLGH 60
Db 23 QSEPEIKLESVVIVSRHGVRAPTKATQIMQDVTPDAMPTWPKIGELTPRGELIAYLGH 82
QY 61 YWRQLVADGLLPKCGCPQSGQGVAIADVDERTKRTGEAFAGLAPDCAITVHTQADTSS 120
Db 83 YWRQLVADGLLPKCGCPQSGQGVAIADVDERTKRTGEAFAGLAPDCAITVHTQADTSS 142
QY 121 PDPLFNPLKTGVCOLDNANVTDAIILRAGGSIADFTGHYQTAFARELERVLANFPQSNCLK 180
Db 143 PDPLFNPLKTGVCOLDNANVTDAIILRAGGSIADFTGHYQTAFARELERVLANFPQSNCLK 202
QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEFGWGRIIDS 240
Db 203 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEFGWGRIIDS 262
QY 241 HOWNILLSLHNAQFDLLQRTPEVARSRATPLLDLTKTALTTHPPKQAYGVTLPTSVLFI 300
Db 263 HOWNILLSLHNAQFVLLQRTPEVARSRATPLLDLTKTALTTHPPKQAYGVTLPTSVLFI 322
QY 301 AGHDTNLANLGALLENLNTLPGQPDNTPPGGELVFERWRRLSDNSQWIQVSLVFOTLQOM 360
Db 323 AGHDTNLANLGALLENLNTLPGQPDNTPPGGELVFERWRRLSDNSQWIQVSLVFOTLQOM 382
QY 361 RDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 410
Db 383 RDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 432
```

Search completed: April 30, 2004, 12:55:03
Job time : 41 secs

GenCore version 5.1.6
Copyright (C) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 30, 2004, 12:36:16 ; Search time 16.4286 Seconds
(without alignments)
1288.405 Million cell updates/sec

Title: US-10-021-723B-13

Perfect score: 2153

Sequence: 1 QSEPELKLESVIVSRHGVR.....CSLAGTQIVNEARIPACSL 410

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/prodata/2/aaa/5A_COMB.pep.*
- 2: /cgn2_6/prodata/2/aaa/5B_COMB.pep.*
- 3: /cgn2_6/prodata/2/aaa/5A_COMB.pep.*
- 4: /cgn2_6/prodata/2/aaa/6B_COMB.pep.*
- 5: /cgn2_6/prodata/2/aaa/PCITUS_COMB.pep.*
- 6: /cgn2_6/prodata/2/aaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2066	96.0	440	3	US-09-259-214-2
2	2066	96.0	440	3	US-09-318-528-2
3	2066	96.0	440	3	US-09-291-931-2
4	2042	94.8	433	4	US-09-540-149A-1
5	1958.5	91.0	423	2	US-08-910-798-2
6	539	25.0	421	4	US-09-489-039A-7512
7	370	17.2	522	4	US-09-489-039A-13501
8	119.5	5.6	467	4	US-09-684-855-165
9	118.5	5.5	439	4	US-09-044-718-3
10	118.5	5.5	440	4	US-09-684-855-106
11	118.5	5.5	440	4	US-09-684-855-128
12	118.5	5.5	440	4	US-09-684-855-151
13	118.5	5.5	449	4	US-09-044-718-12
14	118.5	5.5	465	3	US-08-868-435-33
15	118.5	5.5	465	4	US-08-744-231-33
16	118.5	5.5	465	4	US-09-044-718-78
17	118.5	5.5	465	4	US-09-636-499-6
18	117.5	5.5	465	4	US-09-273-871A-8
19	117.5	5.5	440	4	US-09-684-855-107
20	117.5	5.5	440	4	US-09-684-855-129
21	117.5	5.5	440	4	US-09-684-855-152
22	110.5	5.1	440	4	US-09-684-855-108
23	110.5	5.1	440	4	US-09-684-855-130
24	110.5	5.1	440	4	US-09-684-855-153
25	109.5	5.1	465	4	US-09-044-718-80
26	109.5	5.1	465	4	US-09-636-499-13
27	109.5	5.1	465	4	US-09-044-718-79

28	109.5	5.1	479	1	US-07-923-724-2	Sequence 2, Appli
29	109.5	5.1	479	2	US-08-609-426A-2	Sequence 2, Appli
30	109.5	5.1	479	2	US-08-374-652C-4	Sequence 4, Appli
31	108.5	5.0	440	4	US-09-684-855-109	Sequence 109, App
32	108.5	5.0	440	4	US-09-684-855-131	Sequence 131, App
33	108.5	5.0	440	4	US-09-684-855-154	Sequence 154, App
34	108.5	5.0	455	4	US-09-636-499-14	Sequence 14, Appl
35	108.5	5.0	465	4	US-09-044-718-81	Sequence 81, Appl
36	105	4.9	413	4	US-09-684-855-142	Sequence 142, App
37	105	4.9	422	4	US-09-684-855-119	Sequence 119, App
38	105	4.9	442	3	US-08-993-359-28	Sequence 28, Appl
39	105	4.9	442	4	US-09-273-871A-4	Sequence 4, Appli
40	105	4.9	442	4	US-09-482-558A-28	Sequence 28, Appl
41	103	4.8	1048	4	US-08-887-534A-85	Sequence 85, Appl
42	103	4.8	1048	4	US-09-527-431-85	Sequence 85, Appl
43	101.5	4.7	386	1	US-08-758-213-1	Sequence 1, Appli
44	101.5	4.7	386	2	US-08-692-787-48	Sequence 48, Appl
45	101.5	4.7	386	3	US-09-097-199-48	Sequence 48, Appl

ALIGNMENTS

RESULT 1

US-09-259-214-2

; Sequence 2, Application US/09259214A

; Patent No. 6110719

; GENERAL INFORMATION:

; APPLICANT: Kretz, Keith

; TITLE OF INVENTION: NOVEL PHYTASE

; FILE REFERENCE: DIVER1370-1

; CURRENT APPLICATION NUMBER: US/09/259,214A

; CURRENT FILING DATE: 1999-03-01

; EARLIER APPLICATION NUMBER: 08/910,798

; EARLIER FILING DATE: 1997-08-13

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: PastSeq for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 440

; TYPE: PRT

; ORGANISM: Escherichia coli

US-09-259-214-2

Query Match	96.0%	Score 2066;	DB 3;	Length 440;
Best Local Similarity	97.6%	Pred. No. 2.3e-211;		
Matches 400;	Conservative 0;	Mismatches 10;	Indels 0;	Gaps 0;
QY	1	QSEPELKLESVIVSRHGVRAPTKATQLQMDVTPDAPWTPVKLGELTPRGGLIAYLGH	60	
DB	23	QSEPELKLESVIVSRHGVRAPTKATQLQMDVTPDAPWTPVKLGMLTPRGGLIAYLGH	82	
QY	61	YWRQLVADGLPKCGCPQSGQVAIADVDERTKGTGEAPAGLADPDCAITVHTQADTSS	120	
DB	83	YORQLVADGLLAKKCPQSGQVAIADVDERTKGTGEAPAGLADPDCAITVHTQADTSS	142	
QY	121	PDPLNFKLTGVQCLDNANVTDAILERAGGSADFTGHYQTAFRELERVLPQPSNLCLK	180	
DB	143	PDPLNFKLTGVQCLDNANVTDAILSRAGGSADFTGHYQTAFRELERVLPQPSNLCLK	202	
QY	181	REKQDSCSITQALPSELKVSADCVSLTGAVSLASMLTEIFLQQAQMPGWRITDS	240	
DB	203	REKQDSCSITQALPSELKVSADNVSLTGAVSLASMLTEIFLQQAQMPGWRITDS	262	
QY	241	HWNTLLSLNNAQFDLLQRTPEVARSRATPLDLIKTALTPHPPOKQAYGVTLPTSVLFI	300	
DB	263	HWNTLLSLNNAQFYLLQRTPEVARSRATPLDLINAAALTPHPPOKQAYGVTLPTSVLFI	322	
QY	301	AGHDTNLNAGGALNWTLPQPDNTPPGGELVFRWRRLSDNSQWISVLFQTLQOM	360	
DB	323	AGHDTNLNAGGALNWTLPQPDNTPPGGELVFRWRRLSDNSQWISVLFQTLQOM	382	
QY	361	RDKTFLSLNTPPGVEKLTLAGCEERNAQGCSLAGTQIVNEARIPACSL	410	

Db 383 RDKTFLSNTTPEGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 432

RESULT 2

US-09-318-528-2

; Sequence 2, Application US/09318528

; Patent No. 6183740

; GENERAL INFORMATION:

; APPLICANT: Kretz, Keith

; TITLE OF INVENTION: NOVEL PHYTASE

; FILE REFERENCE: 09010/029003

; CURRENT APPLICATION NUMBER: US/09/318,528

; CURRENT FILING DATE: 1999-05-25

; EARLIER APPLICATION NUMBER: 09/291,931

; EARLIER FILING DATE: 1999-04-13

; EARLIER APPLICATION NUMBER: 08/910,798

; EARLIER FILING DATE: 1997-08-13

; EARLIER APPLICATION NUMBER: 09/259,214

; EARLIER FILING DATE: 1999-03-01

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 440

; TYPE: PRT

; ORGANISM: Escherichia coli

US-09-318-528-2

Query Match 96.0%; Score 2066; DB 3; Length 440;

Best Local Similarity 97.6%; Pred. No. 2.3e-211;

Mismatches 0; Conservative 0; Indels 0; Gaps 0;

Matches 400;

Qy 1 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTTPDAMPWPVKLGELTPRGELIAYLGH 60

Db 23 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTTPDAMPWPVKLGELTPRGELIAYLGH 82

Qy 61 YRQRLVADGLLPKCGCPOSGOVAIADVDETRTKTGEAFAAGLAPDCAITVHTQADTSS 120

Db 83 YRQRLVADGLLPKCGCPOSGOVAIADVDETRTKTGEAFAAGLAPDCAITVHTQADTSS 142

Qy 121 PDPLFNPLKTVGCOLDNANVTDAILLERAGSGIADFTGHYQTAFARELERVLPQSNCLK 180

Db 143 PDPLFNPLKTVGCOLDNANVTDAILLERAGSGIADFTGHYQTAFARELERVLPQSNCLK 202

Qy 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 240

Db 203 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 262

Qy 241 HQWNTLLSHNAQFDLLQRTPEVARSRATPLDLIKTALTTPHPKQAYGVTLPTSVLFI 300

Db 263 HQWNTLLSHNAQFDLLQRTPEVARSRATPLDLIKTALTTPHPKQAYGVTLPTSVLFI 322

Qy 301 AGHDTNLANLGGALELNWTLPGQDNTTPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 360

Db 323 AGHDTNLANLGGALELNWTLPGQDNTTPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 382

Qy 361 RDKTFLSNTTPEGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 410

Db 383 RDKTFLSNTTPEGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 432

RESULT 4

US-09-540-149A-1

; Sequence 1, Application US/09540149A

; Patent No. 6511699

; GENERAL INFORMATION:

; APPLICANT: Lei, Xingen

; TITLE OF INVENTION: ENZYMES WITH IMPROVED PHYTASE ACTIVITY

; FILE REFERENCE: 19603/2791

; CURRENT APPLICATION NUMBER: US/09/540,149A

; CURRENT FILING DATE: 2000-03-31

; PRIOR APPLICATION NUMBER: 60/127,032

; PRIOR FILING DATE: 1999-03-31

; NUMBER OF SEQ ID NOS: 9

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1

; LENGTH: 433

; TYPE: PRT

; ORGANISM: Escherichia coli

; FEATURE:

; NAME/KEY: UNSURE

; LOCATION: (433)

; OTHER INFORMATION: Xaa at position 433 in this sequence is unknown

US-09-540-149A-1

Query Match 94.8%; Score 2042; DB 4; Length 433;

Best Local Similarity 96.6%; Pred. No. 8.1e-209;

Mismatches 396; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Matches 396;

Qy 1 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTTPDAMPWPVKLGELTPRGELIAYLGH 60

Db 23 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTTPDAMPWPVKLGELTPRGELIAYLGH 82

QY 61 YWRQLVADGLLPKCGPQSGQVAIIADVDERTKRTGEAFAAGLAPDCAITVHTQADTSS 120
DB 83 YQRLVADGLLPKCGPQSGQVAIIADVDERTKRTGEAFAAGLAPDCAITVHTQADTSS 142
QY 121 PDPLFNPLKTVGCOLDNANVTDAIIRAGGSIADFTGHYQTAFARELERVLPQSNCLK 180
DB 143 PDPLFNPLKTVGCOLDNANVTDAIIRAGGSIADFTGHYQTAFARELERVLPQSNCLK 202
QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAIVSLASMLTEIFLLQQAQMPGPGWGRITDS 240
DB 203 REKQDESCSLTQALPSELKVSADCVSLTGAIVSLASMLTEIFLLQQAQMPGPGWGRITDS 262
QY 241 HOWNTLLSLHNAQFDLLQRTPEVARSRAATPLDLIKTALTTPHPPOKQAYGVTLPTSVLFI 300
DB 263 HOWNTLLSLHNAQFYLLQRTPEVARSRAATPLDLIKTALTTPHPPOKQAYGVTLPTSVLFI 322
QY 301 AGHDTNLANLGGALELNTLPGQDNTPPGGELVFERWRRLSDNSQWIOVSLVFTLOQM 360
DB 323 AGHDTNLANLGGALELNTLPGQDNTPPGGELVFERWRRLSDNSQWIOVSLVFTLOQM 382
QY 361 RDKTPLSLNTPPGEVKLTLAGCERNNAQMCISLAGFTQIVNEARIPACSL 410
DB 383 RDKTPLSLNTPPGEVKLTLAGCERNNAQMCISLAGFTQIVNEARIPACSL 432

RESULT 5
US-08-910-798-2
; Sequence 2, Application US/08910798
; Patent No. 5876997
; GENERAL INFORMATION:
; APPLICANT: KRETZ
; TITLE OF INVENTION: NOVEL PHYTASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: California
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910,798
; FILING DATE: August 13, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: HAYLE, PH.D., LISA A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09010/029001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 423 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-910-798-2

Query Match 91.0%; Score 1958.5; DB 2; Length 423;
Best Local Similarity 93.4%; Pred. No. 6.2e-200;
Matches 383; Conservative 0; Mismatches 10; Indels 17; Gaps 1;
QY 1 QSEPELKLESVIVSRHGVRAPTKATQLMQDVTDPDAMPVTKLGELTPRGGLIAYLGH 60
DB 23 QSEPELKLESVIVSRHGVRAPTKATQLMQDVTDPDAMPVTKLGELTPRGGLIAYLGH 82

QY 61 YWRQLVADGLLPKCGPQSGQVAIIADVDERTKRTGEAFAAGLAPDCAITVHTQADTSS 120
DB 83 YQRLVADGLLPKCGPQSGQVAIIADVDERTKRTGEAFAAGLAPDCAITVHTQADTSS 125
QY 121 PDPLFNPLKTVGCOLDNANVTDAIIRAGGSIADFTGHYQTAFARELERVLPQSNCLK 180
DB 126 PDPLFNPLKTVGCOLDNANVTDAIIRAGGSIADFTGHYQTAFARELERVLPQSNCLK 185
QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAIVSLASMLTEIFLLQQAQMPGPGWGRITDS 240
DB 186 REKQDESCSLTQALPSELKVSADCVSLTGAIVSLASMLTEIFLLQQAQMPGPGWGRITDS 245
QY 241 HOWNTLLSLHNAQFDLLQRTPEVARSRAATPLDLIKTALTTPHPPOKQAYGVTLPTSVLFI 300
DB 246 HOWNTLLSLHNAQFYLLQRTPEVARSRAATPLDLIKTALTTPHPPOKQAYGVTLPTSVLFI 305
QY 301 AGHDTNLANLGGALELNTLPGQDNTPPGGELVFERWRRLSDNSQWIOVSLVFTLOQM 360
DB 306 AGHDTNLANLGGALELNTLPGQDNTPPGGELVFERWRRLSDNSQWIOVSLVFTLOQM 365
QY 361 RDKTPLSLNTPPGEVKLTLAGCERNNAQMCISLAGFTQIVNEARIPACSL 410
DB 366 RDKTPLSLNTPPGEVKLTLAGCERNNAQMCISLAGFTQIVNEARIPACSL 415

RESULT 6
US-09-489-039A-7512
; Sequence 7512, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 7512
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-7512

Query Match 25.0%; Score 539; DB 4; Length 421;
Best Local Similarity 32.3%; Pred. No. 1.2e-48;
Matches 132; Conservative 70; Mismatches 181; Indels 26; Gaps 9;
QY 1 QSEPE-LKLESVIVSRHGVRAPTKATQLMQDVTDPDAMPVTKLGELTPRGGLIAYL 58
DB 31 KAAPEGYQLQVLLMSRHLRAPLANNQSVLEQSTAKAWPDVGGQLTTKGGVLEVYM 90
QY 59 GHYWRQLVADGLLPKCGPQSGQVAIIADVDERTKRTGEAFAAGLAPDCAITVHTQADT 118
DB 91 GHYWRQLVADGLLPKCGPQSGQVAIIADVDERTKRTGEAFAAGLAPDCAITVHTQADT 150
QY 119 SSPDPLFNPLKTVGCOLDNANVTDAI-IRAGGSIADFTGHYQTAFARELERVLPQSNL 177
DB 151 GTMDPTFNEVITDSDPAFREKALQAMEKROGMQTE-----SYKLETFMIDIRNPS 203
QY 178 CLKREKQDESCSLTQALPSELKVSADCVSLTGAIVSLASMLTEIFLLQQAQMPGPGWGRITDS 235
DB 204 C-----KEKVCVSLSEKDTFSGAGYQCEPGVSGPLKVGNSLVDAFTLQVYEGFPKQVANG 259
QY 236 RITDTSQWNTLLSLHNAQFDLLQRTPEVARSRAATPLDLIKTALTTPHPPOKQAYGVTLPT 295
DB 260 EIADSKQWRVLSKLKNGYQDSLTSTVAQNVAKLVKIDNALVGEAGSK-----A 311
QY 296 SVLFTAGHDTNLANLGGALELNTLPGQDNTPPGGELVFERWRRLSDNSQWIOVSLVFT 354
DB 312 KVTLLVGHDSNTIASLTLTALDFPKYQLPQGYERTPIGGKLLFORWHDSSAGNRDLMKIEYV 371

QY 3 EPELKES-----VVIVSRHGVRAPT-----KATQLMQDVTDPD-----WP 38
DB 31 EDELSVSSKLPKDCRITLVQVLSRHGARYPTSSKSKYKLVTAIQANATDFKGFALFK 90
QY 39 TWPVKLG--ELTPRGELIAYLG--HYWRQRLVADGLLPKCGCPSQGVVAIIADVDERT 94
DB 91 TYNVTLGADDLTPFGEOQLVNSGKIFYORYKALARSVVP-----FIRASGSDRVI 140
QY 95 KTGEAFAAGL-----APDCAITVHTQADTSSPDPLFNLKTVGC-OLDNAN 139
DB 141 ASGEKFIQFOQAKLADPGATNRAAPAI SVII--PESETFN-----NTLDHGVCITKFEASQ 194
QY 140 VTDAILERAGGSIADFTGHYQTAFRELVRVLPFQSNCLCKREKQDESCSLTQALPSELK 199
DB 195 LGDEV-----AANFTALFAPDIR-----ARAEKH-----LPGVTL 224
QY 200 VSADCVSLTGAVSLASMLTEIFLLQQAQMPGPGWGRITDHSQWNTLLSL--HN--AQFD 255
DB 225 TDEDVVSMDMCSFDTV-----ARTSDASQLSPFCQLFTHNEWKYN 266
QY 256 LLQR-----TPE-VARSRATPLLDLIKT--ALTPHPQKQAYGV 291
DB 267 YLQSLGKYGYGAGNPLGPAQGIQFTNELIARLTRSPVQDHTSTNSTLVSNP-----A 319
QY 292 TLPTSVLFTA--GHDNTNLNGLGALNLTLPQPDNTPPGGELVFERWRRLSD-NSQWI 348
DB 320 TFFLNATMYVDVDFSHDMSVSIFFAL-----GLYNGTEPLSRTSVESAKELDGYASWV 372
QY 349 ---QVSLVFTLOQMOKDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEAR 404
DB 373 VPFGARAYFETMCKSEKEPL-VRALINDRVVPLHGCDV-DKLGRCCKLNDVFKGLSWAR 429

RESULT 10
US-09-684-855-106
; Sequence 106, Application US/09684855
; Patent No. 6599735
; GENERAL INFORMATION:
; APPLICANT: P. Hoffmann-La Roche AG
; TITLE OF INVENTION: CONTINUOUS FERMENTATION PROCESS
; FILE REFERENCE: C38435/111692
; CURRENT APPLICATION NUMBER: US/09/684,855
; CURRENT FILING DATE: 2000-10-06
; PRIOR FILING DATE: 2000-10-04
; PRIOR FILING DATE: 2000-10-04
; PRIOR FILING DATE: 1999-10-11
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 106
; LENGTH: 440
; TYPE: PRT
; ORGANISM: A. fumigatus 13073
US-09-684-855-106

Query Match 5.5%; Score 118.5; DB 4; Length 440;
Best Local Similarity 20.9%; Pred. No. 0.00088;
Matches 100; Conservative 65; Mismatches 157; Indels 157; Gaps 26;
QY 3 EPELKES-----VVIVSRHGVRAPT-----KATQLMQDVTDPD-----WP 38
DB 32 EDELSVSSKLPKDCRITLVQVLSRHGARYPTSSKSKYKLVTAIQANATDFKGFALFK 91
QY 39 TWPVKLG--ELTPRGELIAYLG--HYWRQRLVADGLLPKCGCPSQGVVAIIADVDERT 94
DB 92 TYNVTLGADDLTPFGEOQLVNSGKIFYORYKALARSVVP-----FIRASGSDRVI 141
QY 95 KTGEAFAAGL-----APDCAITVHTQADTSSPDPLFNLKTVGC-OLDNAN 139
DB 142 ASGEKFIQFOQAKLADPGATNRAAPAI SVII--PESETFN-----NTLDHGVCITKFEASQ 195
QY 140 VTDAILERAGGSIADFTGHYQTAFRELVRVLPFQSNCLCKREKQDESCSLTQALPSELK 199
DB 195 LGDEV-----AANFTALFAPDIR-----ARAEKH-----LPGVTL 225
QY 200 VSADCVSLTGAVSLASMLTEIFLLQQAQMPGPGWGRITDHSQWNTLLSL--HN--AQFD 255
DB 226 TDEDVVSMDMCSFDTV-----ARTSDASQLSPFCQLFTHNEWKYN 267
QY 256 LLQR-----TPE-VARSRATPLLDLIKT--ALTPHPQKQAYGV 291
DB 268 YLQSLGKYGYGAGNPLGPAQGIQFTNELIARLTRSPVQDHTSTNSTLVSNP-----A 320
QY 292 TLPTSVLFTA--GHDNTNLNGLGALNLTLPQPDNTPPGGELVFERWRRLSD-NSQWI 348
DB 321 TFFLNATMYVDVDFSHDMSVSIFFAL-----GLYNGTEPLSRTSVESAKELDGYASWV 373

DB 196 LGDEV-----AANFTALFAPDIR-----ARAEKH-----LPGVTL 225
QY 200 VSADCVSLTGAVSLASMLTEIFLLQQAQMPGPGWGRITDHSQWNTLLSL--HN--AQFD 255
DB 226 TDEDVVSMDMCSFDTV-----ARTSDASQLSPFCQLFTHNEWKYN 267
QY 256 LLQR-----TPE-VARSRATPLLDLIKT--ALTPHPQKQAYGV 291
DB 268 YLQSLGKYGYGAGNPLGPAQGIQFTNELIARLTRSPVQDHTSTNSTLVSNP-----A 320
QY 292 TLPTSVLFTA--GHDNTNLNGLGALNLTLPQPDNTPPGGELVFERWRRLSD-NSQWI 348
DB 321 TFFLNATMYVDVDFSHDMSVSIFFAL-----GLYNGTEPLSRTSVESAKELDGYASWV 373
QY 349 ---QVSLVFTLOQMOKDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEAR 404
DB 374 VPFGARAYFETMCKSEKEPL-VRALINDRVVPLHGCDV-DKLGRCCKLNDVFKGLSWAR 430

RESULT 11
US-09-684-855-128
; Sequence 128, Application US/09684855
; Patent No. 6599735
; GENERAL INFORMATION:
; APPLICANT: P. Hoffmann-La Roche AG
; TITLE OF INVENTION: CONTINUOUS FERMENTATION PROCESS
; FILE REFERENCE: C38435/111692
; CURRENT APPLICATION NUMBER: US/09/684,855
; CURRENT FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: EP 00121663.9
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: EP 99120289.6
; PRIOR FILING DATE: 1999-10-11
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 128
; LENGTH: 440
; TYPE: PRT
; ORGANISM: A. fumigatus 13073
US-09-684-855-128

Query Match 5.5%; Score 118.5; DB 4; Length 440;
Best Local Similarity 20.9%; Pred. No. 0.00088;
Matches 100; Conservative 65; Mismatches 157; Indels 157; Gaps 26;
QY 3 EPELKES-----VVIVSRHGVRAPT-----KATQLMQDVTDPD-----WP 38
DB 32 EDELSVSSKLPKDCRITLVQVLSRHGARYPTSSKSKYKLVTAIQANATDFKGFALFK 91
QY 39 TWPVKLG--ELTPRGELIAYLG--HYWRQRLVADGLLPKCGCPSQGVVAIIADVDERT 94
DB 92 TYNVTLGADDLTPFGEOQLVNSGKIFYORYKALARSVVP-----FIRASGSDRVI 141
QY 95 KTGEAFAAGL-----APDCAITVHTQADTSSPDPLFNLKTVGC-OLDNAN 139
DB 142 ASGEKFIQFOQAKLADPGATNRAAPAI SVII--PESETFN-----NTLDHGVCITKFEASQ 195
QY 140 VTDAILERAGGSIADFTGHYQTAFRELVRVLPFQSNCLCKREKQDESCSLTQALPSELK 199
DB 196 LGDEV-----AANFTALFAPDIR-----ARAEKH-----LPGVTL 225
QY 200 VSADCVSLTGAVSLASMLTEIFLLQQAQMPGPGWGRITDHSQWNTLLSL--HN--AQFD 255
DB 226 TDEDVVSMDMCSFDTV-----ARTSDASQLSPFCQLFTHNEWKYN 267
QY 256 LLQR-----TPE-VARSRATPLLDLIKT--ALTPHPQKQAYGV 291
DB 268 YLQSLGKYGYGAGNPLGPAQGIQFTNELIARLTRSPVQDHTSTNSTLVSNP-----A 320
QY 292 TLPTSVLFTA--GHDNTNLNGLGALNLTLPQPDNTPPGGELVFERWRRLSD-NSQWI 348
DB 321 TFFLNATMYVDVDFSHDMSVSIFFAL-----GLYNGTEPLSRTSVESAKELDGYASWV 373

Qy 349 ---QVSLVFOTLQMRDKTFLSLNTPPEVKLTLAGCEERNAQCMGSLAGFTQIVNEAR 404
Db 374 VPFGARAYFETMCKSEKPL-VRALINDRVVPLHGCDV-DKLGRCCKLNDVFKGLSWAR 430

RESULT 12

US-09-684-855-151

; Sequence 151, Application US/09684855

; Patent No. 6599735

; GENERAL INFORMATION:

; APPLICANT: F. Hoffmann-La Roche AG

; TITLE OF INVENTION: CONTINUOUS FERMENTATION PROCESS

; FILE REFERENCE: C38435/111692

; CURRENT APPLICATION NUMBER: US/09/684,855

; CURRENT FILING DATE: 2000-10-06

; PRIOR APPLICATION NUMBER: EP 00121663.9

; PRIOR FILING DATE: 2000-10-04

; PRIOR APPLICATION NUMBER: EP 99120289.6

; PRIOR FILING DATE: 1999-10-11

; NUMBER OF SEQ ID NOS: 169

; SOFTWARE: Patent in Ver. 2.0

; SEQ ID NO 151

; LENGTH: 440

; TYPE: PRT

; ORGANISM: A. fumigatus 32722

US-09-684-855-151

Query Match 5.5%; Score 118.5; DB 4; Length 440;
Best Local Similarity 20.9%; Pred. No. 0.00088;
Matches 100; Conservative 65; Mismatches 157; Indels 157; Gaps 26;

Qy 3 EPELKLES-----VVIVSRHGVRAPT-----KATQMODVTPDA-----WP 38
Db 32 EDELSVSKLPKDCRITLVOVLSRHGARYPTSSKSKYKLVTAIQANATDKGKPAFLK 91
Qy 39 TWPKVIG--ELTPRGGLIAYLG--HYWRORLVADGLPKCGCPQSGQVAILIADVDERT 94
Db 92 TYNVTLGADDLTPFGEOQLVNSGKIFYQRYKALARSVVP-----FIRASGSDRVI 141
Qy 95 KTGEAFAAGL-----APDCAITVHTQADTSSPDLPFLNPLKTGVC--QLDNAN 139
Db 142 ASGEKFTGEGQOAKLADPGATNRAAPAIISVII--PESETFN-----NTLDHGVCCTKFEASQ 195
Qy 140 VTDAILERAGGSIAADFTGHYQTAFRELERVLPQSNCLKREKQDESCSLTQALPSELK 199
Db 196 LGDEV-----AANFTALFAPDIR-----ARAEKH-----LPGVTL 225
Qy 200 VSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDSHOWNTLLSL--HN--AQFD 255
Db 226 TDEDVWSLMDMCSFDTV-----ARTSDASQLSPFCQLFTHNEWKYN 267
Qy 256 LLQR-----TPE-VARSRATPLLDLIKT--ALTPHPPOKQAYGV 291
Db 268 YLOSGLKYYGYGAGNPLGPAQGGIGFTNELIARLTRSPVDHTSTNSTLVSNP-----A 320
Qy 292 TLPTSVLVFLA--GHDTNLANLGALNELNWLTPGPDNTPPGGELVFRWRRLSD--NSQWI 348
Db 321 TFFPLNATMYVDVFDHNSMVSIFPAL-----GLYNGTEPLSRTSVESAKELDGYASVW 373
Qy 349 ---QVSLVFOTLQMRDKTFLSLNTPPEVKLTLAGCEERNAQCMGSLAGFTQIVNEAR 404
Db 374 VPFGARAYFETMCKSEKPL-VRALINDRVVPLHGCDV-DKLGRCCKLNDVFKGLSWAR 430

RESULT 13

US-09-044-718-12

; Sequence 12, Application US/09044718

; Patent No. 6391605

; GENERAL INFORMATION:

; APPLICANT: KOSTREWA, Dirk

; APPLICANT: PASAMONTES, Luis

; APPLICANT: TOMSCHY, Andrea

; APPLICANT: van LOON, Adolphus

APPLICANT: VOGEL, Kurt
APPLICANT: WYSS, Markus
TITLE OF INVENTION: MODIFIED PHYTASES
FILE REFERENCE: Modified Phytases
CURRENT APPLICATION NUMBER: US/09/044,718
CURRENT FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: EP 97810175.6
PRIOR FILING DATE: 1997-03-25
NUMBER OF SEQ ID NOS: 82
SOFTWARE: Patent in Ver. 2.1
SEQ ID NO 12
LENGTH: 449
TYPE: PRT
ORGANISM: Aspergillus fumigatus
US-09-044-718-12

Query Match 5.5%; Score 118.5; DB 4; Length 449;
Best Local Similarity 20.9%; Pred. No. 0.00091;
Matches 100; Conservative 65; Mismatches 157; Indels 157; Gaps 26;

Qy 3 EPELKLES-----VVIVSRHGVRAPT-----KATQMODVTPDA-----WP 38
Db 41 EDELSVSKLPKDCRITLVOVLSRHGARYPTSSKSKYKLVTAIQANATDKGKPAFLK 100
Qy 39 TWPKVIG--ELTPRGGLIAYLG--HYWRORLVADGLPKCGCPQSGQVAILIADVDERT 94
Db 101 TYNVTLGADDLTPFGEOQLVNSGKIFYQRYKALARSVVP-----FIRASGSDRVI 150
Qy 95 KTGEAFAAGL-----APDCAITVHTQADTSSPDLPFLNPLKTGVC--QLDNAN 139
Db 151 ASGEKFTGEGQOAKLADPGATNRAAPAIISVII--PESETFN-----NTLDHGVCCTKFEASQ 204
Qy 140 VTDAILERAGGSIAADFTGHYQTAFRELERVLPQSNCLKREKQDESCSLTQALPSELK 199
Db 205 LGDEV-----AANFTALFAPDIR-----ARAEKH-----LPGVTL 234
Qy 200 VSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDSHOWNTLLSL--HN--AQFD 255
Db 235 TDEDVWSLMDMCSFDTV-----ARTSDASQLSPFCQLFTHNEWKYN 276
Qy 256 LLQR-----TPE-VARSRATPLLDLIKT--ALTPHPPOKQAYGV 291
Db 277 YLOSGLKYYGYGAGNPLGPAQGGIGFTNELIARLTRSPVDHTSTNSTLVSNP-----A 329
Qy 292 TLPTSVLVFLA--GHDTNLANLGALNELNWLTPGPDNTPPGGELVFRWRRLSD--NSQWI 348
Db 330 TFFPLNATMYVDVFDHNSMVSIFPAL-----GLYNGTEPLSRTSVESAKELDGYASVW 382
Qy 349 ---QVSLVFOTLQMRDKTFLSLNTPPEVKLTLAGCEERNAQCMGSLAGFTQIVNEAR 404
Db 383 VPFGARAYFETMCKSEKPL-VRALINDRVVPLHGCDV-DKLGRCCKLNDVFKGLSWAR 439

RESULT 14

US-08-868-435-33

; Sequence 33, Application US/08868435

; Patent No. 6291221

; GENERAL INFORMATION:

; APPLICANT: Van Loon, Adolphus

; APPLICANT: Mitchell, David

; TITLE OF INVENTION: POLYPEPTIDES WITH PHYTASE ACTIVITY

; NUMBER OF SEQUENCES: 35

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Hoffmann-La Roche Inc.

; STREET: 340 Kingsland Street

; CITY: Nutley

; STATE: New Jersey

; COUNTRY: United States of America

; ZIP: 07110

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/868,435
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/744,231
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Kass, Alan P
REGISTRATION NUMBER: 32142
REFERENCE/DOCKET NUMBER: Case Docket 9339
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-4205
TELEFAX: (201) 235-2363
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 465 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: misc_feature
LOCATION: 104
OTHER INFORMATION: /note="potential N-glycosylation site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 119
OTHER INFORMATION: /note="potential N-glycosylation site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 205
OTHER INFORMATION: /note="potential N-glycosylation site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 228
OTHER INFORMATION: /note="potential N-glycosylation site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 337
OTHER INFORMATION: /note="potential N-glycosylation site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 374
OTHER INFORMATION: /note="potential N-glycosylation site"
US-08-868-435-33

Query Match 5.5%; Score 118.5; DB 3; Length 465;
Best Local Similarity 20.9%; Pred. No. 0.00096;
Matches 100; Conservative 65; Mismatches 157; Indels 157; Gaps 26;

Qy 3 EPELKLES-----VVIVSRHGVRAPT-----KATQLMDQVTPDA-----WP 38
Db 57 EDELSVSKLPKDCRITLVQVLSHGARYPTSSKSKYKLVTAIQANATDPKGFAPLK 116
Qy 39 TWPVKLG--ELTPRGGLIAYLG--HWYRORLVADGLLPKCGCPQSQOVAIADVBERTR 94
Db 117 TYNLTGADDLTPFGEOQLVNSGKIFYQRYKALARSVP-----FIRASGSDRVI 166
Qy 95 KTGEAFPAAGL-----APCAITVHTQADTSSPDPPLFNPLKTVGC-OLDNAN 139
Db 167 ASGKEPIEGFOQAKLADPGATNRAPALSVII-PESETFN-----NVLHDGVCTKFEASQ 220
Qy 140 VTDAILERAGGSIADFTGHVQTAFRELERVNFQSNLCUKRKQKQESCLTQALPSELK 199
Db 221 LGDEV-----AANFTALEFAPDIR-----ARAERH-----LPGVTL 250
Qy 200 VSADCVSITGVASLASMLTEIFLLQQAQMGPEPCMGRIITSHOWNTLLSL--HN--AQFD 255
Db 251 TDEDVSLMDMCSFDTV-----ARTSDASQLSPFCQLFTHEWKKYN 292
Qy 256 LLQR-----TPE-VARSATPFLDLIKT--ALTPHPQKQAYGV 291

Db 293 YLQSLGKYGYGAGNPLGPAQGIQFTNELIARLTRSPVQDHTSTNSTLVSNP-----A 345
Qy 292 TLPTSVLFFIA--GHDITNLANLGGALELNWTLPGQPDNTPPGGELVFERWRRLSD-NSQWI 348
Db 346 TFFLNATMYVDFSHDSNMVSIFFAL-----GLYNGTEPLSRTSVESAKELDGYASASWV 398
Qy 349 ---QVSLVFQTLQOMRDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEAR 404
Db 399 VPFGARAYFETMOCKSEKPL-VREALINDRVVPLHGCDV-DKLGRCKLINDFVKGLSWAR 455
RESULT 15
US-08-744-231-33
Sequence 33, Application US/08744231
Patent No. 6358722
GENERAL INFORMATION:
APPLICANT: Van Loon, Adolphus
APPLICANT: Mitchell, David
TITLE OF INVENTION: POLYPEPTIDES WITH PHYTASE ACTIVITY
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: New Jersey
COUNTRY: United States of America
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/744,231
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/424,757
FILING DATE: 18-APR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Kass, Alan P
REGISTRATION NUMBER: 32142
REFERENCE/DOCKET NUMBER: Case Docket 9339
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-4205
TELEFAX: (201) 235-2363
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 465 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: misc_feature
LOCATION: 104
OTHER INFORMATION: /note="potential N-glycosylation site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 119
OTHER INFORMATION: /note="potential N-glycosylation site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 205
OTHER INFORMATION: /note="potential N-glycosylation site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 228
OTHER INFORMATION: /note="potential N-glycosylation site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 337
OTHER INFORMATION: /note="potential N-glycosylation site"
FEATURE:

```
; NAME/KEY: misc_feature
; LOCATION: 374
; OTHER INFORMATION: /note="potential N-glycosylation site"
US-08-744-231-33

Query Match      5.5%; Score 118.5; DB 4; Length 465;
Best Local Similarity 20.9%; Pred. No. 0.00096;
Matches 100; Conservative 65; Mismatches 157; Indels 157; Gaps 26;

QY 3 EPELKLES-----VVVSRHGVRAPT-----KATQLMQDVTTPDA-----WP 38
Db 57 EDELSVSSKLPKDCRITLVQVLSRHGARYPTSSKKYKLVTAIQANATDFKGFAPLAK 116
QY 39 TWPVKLG--ELTPRGGELIAYLG--HYWRQLVADGLLPKCGPQSGQVAILIADVDERTR 94
Db 117 TYNVTLGADDLTFPGQLVNSGIKFYQYKALARSVP-----FIRASGSDAVI 166
QY 95 KTGEEAFAAGL-----APDCAITVHTQADTSSPDPLFNPLKTGVC-QLDNAN 139
Db 167 ASGEKFIQGFQQAQKADPGATNRAAPAIISVII-PESETFN-----NTLDHGVCTKFEASQ 220
QY 140 VTDAILERAGGSIADEFTGHYQTAFRELERVLNFPQSNLCLKREKQDESCSLTOALPSELK 199
Db 221 LGDEV-----AANTALFAPDIR-----ARAEKH-----LPGVTL 250
QY 200 VSADCVSLTGAVSLASMLTEIFILQQAQGMPEPGWGRITDSHOWNTLLSL--HN--AOPD 255
Db 251 TDEDVVSMDMCSFDTV-----ARTSDASQSPFCQLFTHNEWKKYN 292
QY 256 LLQR-----TPE-VARSRAATPLADLIKT--ALTPHPQKQAYGV 291
Db 293 YLOSGLKYYGYGAGNPLGPAQGIPTNELIARLTRSPVDHTSTNSTLVSNP-----A 345
QY 292 TLPTSVLFIA--CHDTNLANLGGALELNTLPQPDNTPPGGELVFERWRRLSD-NSQWI 348
Db 346 TFEPLNATMYVDFSHDNSMWSIFPAL-----GLYNGTEPLSRTSVESAKELDGYASWV 398
QY 349 ---QVSLVFOTLQOMRDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEAR 404
Db 399 VPFGARAYPETMQCKSEKEPL-VRALINDRVVPLHGCDV-DKLGRCCKLNDFVKGLSWAR 455
```

Search completed: April 30, 2004, 12:43:17
Job time : 17.4286 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 30, 2004, 12:30:35 ; Search time 52.5 Seconds

(without alignments)
2206.561 Million cell updates/sec

Title: US-10-021-723B-13

Perfect score: 2153

Sequence: 1 QSEPEKLSEVIVSRHGVR.....CSLAGFTQIVNEARIPACSL 410

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_29Jan04:*

- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001s:*
- 5: Geneseq2002s:*
- 6: Geneseq2003as:*
- 7: Geneseq2003bs:*
- 8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2153	100.0	410	5	ABP51937 Phytase p
2	2153	100.0	412	6	AAE37851 Maize-opt
3	2153	100.0	412	7	ABR82310 Nov9x phy
4	2153	100.0	431	6	AAE37853 pNOV4054
5	2153	100.0	436	7	ADC87743 Modified
6	2153	100.0	437	6	AAE37854 pNOV4058
7	2123	98.6	430	5	AAE15808 Escherich
8	2123	98.6	430	6	ADA19451 E. coli K
9	2077	96.5	432	3	AAE36257 Lama2/APP
10	2077	96.5	432	3	AAE36261 R15/APP
11	2077	96.5	432	3	AAE36262 SV40/APP
12	2077	96.5	432	3	AAE36259 R15/APP
13	2077	96.5	432	3	AAE36258 R15/APP
14	2077	96.5	432	3	AAE36263 Lama2/APP
15	2077	96.5	432	3	AAE36260 R15/APP
16	2077	96.5	432	3	AAU77755 Phytase a
17	2077	96.5	432	5	AAE15807 Escherich
18	2077	96.5	432	5	ABP3625 Acid phos
19	2077	96.5	432	6	ADA19450 E. coli K
20	2077	96.5	432	6	ABU15427 Protein e
21	2077	96.5	432	7	ADC87745 Escherich
22	2072	96.2	432	4	AAE02631 E. coli a
23	2066	96.0	432	4	AAE02634 E. coli a
24	2066	96.0	440	4	AAE37892 Escherich
25	2066	96.0	440	5	AAE22836 Escherich

ALIGNMENTS

RESULT 1

ABP51937
ID ABP51937 standard; protein; 410 AA.

XX AC ABP51937;

XX DT 08-OCT-2002 (first entry)

XX DE Phytase protein sequence.

XX KW Phytase; enzyme; E.C.3.1.3.8; phytate; foodstuff; feed.

XX OS Unidentified.

XX PN WO200248332-A2.

XX PD 20-JUN-2002.

XX PF 12-DEC-2001; 2001WO-US048774.

XX PR 12-DEC-2000; 2000US-0255090P.

XX PA (DIVE-) DIVERSA CORP.

XX PI Short J, Mathur EJ, Richardson T, Robertson D, Barton N;

XX DR WPI; 2002-583504/62.

XX PT Novel recombinant phytase protein and polynucleotide for improving

XX PT nutritional value of phytate-containing foodstuff, in animal feed and

XX PT feed supplements and to degrade excess phytase from environment or

XX PT sample.

XX PS Disclosure; Fig 9; 208pp; English.

The present invention describes an isolated phytase protein (I). (I) can be used for improving the nutritional value of a phytate-containing foodstuff, by contacting the phytate-containing foodstuff with (I), where the enzyme catalyses the liberation of inorganic phosphate from the phytate-containing foodstuff, and so improving the nutritive value of the contacted foodstuff. The liberation of inorganic phosphate occurs prior to or after the ingestion of phytate-containing foodstuff by a recipient organism. Nucleotide sequences (II) encoding (I) can be used for producing an animal feed, by transforming a plant, plant portion or plant cell with a nucleic acid expression vector, comprising (II), culturing the plant, plant portion or plant cell under conditions in which the phytase protein is expressed and converting the plant, plant portion or plant cell into a composition suitable for animal feed. The animal is

CC preferably a monogastric animal or a ruminant. (I) and (II) are useful in
 CC animal feed and feed supplements as well as in treatments to degrade or
 CC remove excess phytate from the environment or a sample. (I) reduces
 CC phytate levels in animal manure and so reduces the phosphate pollution of
 CC the environment. The present sequence represents a phytase from the
 CC present invention
 XX
 XX

SQ Sequence 410 AA;
 Query Match 100.0%; Score 2153; DB 5; Length 410;
 Best Local Similarity 100.0%; Pred. No. 1.4e-210;
 Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 QSEPELKLESVIVSRHGVRAPTKATQMDVTPDAMPVTKLGGELIAYLGH 60
 DB 1 QSEPELKLESVIVSRHGVRAPTKATQMDVTPDAMPVTKLGGELIAYLGH 60
 QY 61 YWQRLVADGLLPKCGPQSGQVVAIIADVDERTKGTGEAFAAGLAPDCAITVHTQADTSS 120
 DB 61 YWQRLVADGLLPKCGPQSGQVVAIIADVDERTKGTGEAFAAGLAPDCAITVHTQADTSS 120
 QY 121 PDPLFNPLKTVGVCQOLDNANVTDAILERAGGSIADFTGHYQTAFRELERVLPQSNCLK 180
 DB 121 PDPLFNPLKTVGVCQOLDNANVTDAILERAGGSIADFTGHYQTAFRELERVLPQSNCLK 180
 QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFILQOAGMPEPGWGRITDS 240
 DB 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFILQOAGMPEPGWGRITDS 240
 QY 241 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLLDLIKTALTTPHPPOKQAYGTVLPTSVLFI 300
 DB 241 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLLDLIKTALTTPHPPOKQAYGTVLPTSVLFI 300
 QY 301 AGHDTNLANLGGALELNTWLPQGPDPNTPPGGELVFERWRRLSDNSQWIOVSLVFTQLQOM 360
 DB 301 AGHDTNLANLGGALELNTWLPQGPDPNTPPGGELVFERWRRLSDNSQWIOVSLVFTQLQOM 360
 QY 361 RDKTPLSLNTPPGGVKLTLAGCERNAGMCSLAGFTQIVNEARIPACSL 410
 DB 361 RDKTPLSLNTPPGGVKLTLAGCERNAGMCSLAGFTQIVNEARIPACSL 410

RESULT 2
 AAE37851
 ID AAE37851 standard; protein; 412 AA.
 XX
 AC AAE37851;
 XX
 DT 06-NOV-2003 (first entry)
 XX
 DE Maize-optimised Nov9X phytase mutant.
 XX
 KW Thermotolerant phytase; weight gain; animals fed diet; grain processing;
 KW nutritive value; transformed plant; anabolic; maize; mutant; mutain.
 XX
 OS Zea mays.
 OS Synthetic.
 XX
 PN WO2003057248-A1.
 XX
 PD 17-JUL-2003.
 XX
 PF 30-DEC-2002; 2002WO-US041787.
 XX
 PR 28-DEC-2001; 2001US-0344476P.
 XX
 PA (SYGN) SYNGENTA PARTICIPATIONS AG.
 XX
 PI Lanahan MB, Betts S;
 XX
 DR WPI; 2003-607980/57.
 DR N-PSDB; AAD57147.
 XX

PT Preparing a thermotolerant phytase for preparing animal feed or human
 PT food by expressing in a plant cell an expression cassette comprising a
 PT promoter operably linked to a nucleic acid molecule encoding a
 XX thermotolerant phytase.
 XX

Claim 3; Page 89; 157pp; English.
 XX
 CC The invention relates to a method for preparing a thermotolerant phytase.
 CC The method comprises expressing in a plant cell an expression cassette
 CC comprising a promoter operably linked to a nucleic acid molecule encoding
 CC a thermotolerant phytase which retains at least 40% activity after 30
 CC minutes at 60plusOC and has a specific activity of greater than 200 U/mg
 CC at pH 4.5 and 37plusOC. The method is useful for preparing a
 CC thermotolerant phytase for preparing animal feed or human food. The
 CC invention is useful for reducing the feed conversion ratio and increasing
 CC weight gain, improving reducing feed conversion ratios or increasing
 CC weight gain of animals fed diets with inorganic phosphate at levels below
 CC 0.45%, minimising dietary requirements of phosphorus in an animal,
 CC enhancing the utilisation of phosphorus present in animal feed, enhancing
 CC organic phosphorus utilisation from organic phosphorus sources in animal
 CC feed, decreasing the phosphate levels in excreta from an animal,
 CC improving the processing of grain, improving the nutritive value of
 CC processed grain product or a method of processing grain, improving the
 CC nutritive value of animal feed and human food, and preparing a
 CC transformed plant which expresses a thermotolerant phytase. The present
 CC sequence is maize-optimised Nov9X phytase mutant used in the
 CC exemplification of the invention
 XX

SQ Sequence 412 AA;
 Query Match 100.0%; Score 2153; DB 6; Length 412;
 Best Local Similarity 100.0%; Pred. No. 1.4e-210;
 Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 QSEPELKLESVIVSRHGVRAPTKATQMDVTPDAMPVTKLGGELIAYLGH 60
 DB 3 QSEPELKLESVIVSRHGVRAPTKATQMDVTPDAMPVTKLGGELIAYLGH 62
 QY 61 YWQRLVADGLLPKCGPQSGQVVAIIADVDERTKGTGEAFAAGLAPDCAITVHTQADTSS 120
 DB 63 YWQRLVADGLLPKCGPQSGQVVAIIADVDERTKGTGEAFAAGLAPDCAITVHTQADTSS 122
 QY 121 PDPLFNPLKTVGVCQOLDNANVTDAILERAGGSIADFTGHYQTAFRELERVLPQSNCLK 180
 DB 123 PDPLFNPLKTVGVCQOLDNANVTDAILERAGGSIADFTGHYQTAFRELERVLPQSNCLK 182
 QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFILQOAGMPEPGWGRITDS 240
 DB 183 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFILQOAGMPEPGWGRITDS 242
 QY 241 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLLDLIKTALTTPHPPOKQAYGTVLPTSVLFI 300
 DB 243 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLLDLIKTALTTPHPPOKQAYGTVLPTSVLFI 302
 QY 301 AGHDTNLANLGGALELNTWLPQGPDPNTPPGGELVFERWRRLSDNSQWIOVSLVFTQLQOM 360
 DB 303 AGHDTNLANLGGALELNTWLPQGPDPNTPPGGELVFERWRRLSDNSQWIOVSLVFTQLQOM 362
 QY 361 RDKTPLSLNTPPGGVKLTLAGCERNAGMCSLAGFTQIVNEARIPACSL 410
 DB 363 RDKTPLSLNTPPGGVKLTLAGCERNAGMCSLAGFTQIVNEARIPACSL 412

RESULT 3
 ABR82310
 ID ABR82310 standard; protein; 412 AA.
 XX
 AC ABR82310;
 XX
 DT 06-NOV-2003 (first entry)
 XX
 DE Nov9X phytase polypeptide.
 XX

Nov9X; thermotolerant; phytase; phosphate; animal food; phosphorous; enzyme.
 Synthetic.
 Key Location/Qualifiers
 Misc-difference 1 /note= "encoded by GAG"
 W02003057247-A1.
 17-JUL-2003.
 30-DEC-2002; 2002WO-US041785.
 28-DEC-2001; 2001US-0344523P.
 (SYGN) SYNGENTA PARTICIPATIONS AG.
 (DIVE-) DIVERSA CORP.
 Lanahan ML, Koepf E, Kretz K;
 WPI; 2003-598328/56.
 N-PSDB; ACF45789.
 Preparing thermotolerant phytase useful for preparing animal feed or human food, by expressing in microbial host cell expression cassette comprising a promoter operably linked to a nucleic acid encoding the enzyme.
 Claim 11; Fig 1B; 79pp; English.
 The invention relates to preparing a thermotolerant phytase (I), involving expressing in a microbial host cell an expression cassette comprising a promoter operably linked to a nucleic acid molecule encoding (I). The thermotolerant phytase has a specific activity of greater than 400 U/mg, 600 U/mg or 800 U/mg at pH 4.5 and 37plusOC. The nucleic acid molecule encodes a fusion polypeptide comprising the thermotolerant phytase and a signal sequence which is operably linked to the thermotolerant phytase. The thermotolerant phytase is glycosylated and has a simulated gastric half-life of greater than 25 minutes at a pH greater than 2.0 and less than 4.0. (I) is useful for preparing animal feed, improving nutritive value of animal feed and human food, and for preparing human food. A heat-treated animal feed mixture comprising an inorganic phosphate at below 0.45% and (I) is useful for reducing feed conversion ratios or increasing weight gain of animals fed diets with inorganic phosphate at levels below 0.45%. It is also useful for decreasing phosphate levels in excreta from an animal. An animal feed composition comprising (I) is useful for decreasing the feed conversion ratio and increasing the weight gain of an animal. The composition is useful for minimizing inorganic dietary requirements of phosphorous in an animal, for enhancing utilization of phosphorous present in feed for an animal, and for increasing organic phosphate utilization from organic phosphorus sources in feed for an animal, where the feed increases the bioavailability of inorganic phosphorus in the feed to the animal. It is also useful for decreasing phosphate levels in excreta from an animal. In all the above mentioned conditions, the feed is a poultry or swine feed. The thermotolerant phytase has a half-life of about 30 minutes in the digestive tract of the animal. The present sequence represents a thermotolerant Nov9X phytase
 Sequence 412 AA;
 Query Match 100.0%; Score 2153; DB 7; Length 412;
 Best Local Similarity 100.0%; Pred. No. 1.4e-210;
 Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 1 QSEPELKLESVIVSRHGVRAPTATQLMQDVTDPAMPTWPKLGELTPRGELIAYLGH 60
 3 QSEPELKLESVIVSRHGVRAPTATQLMQDVTDPAMPTWPKLGELTPRGELIAYLGH 62
 51 YWRQLVADGLLPKCGPQSGQVAIIADVDETRKTGEAFAGLAPDCAITVHTQADTSS 120

Db 63 YWRQLVADGLLPKCGPQSGQVAIIADVDETRKTGEAFAGLAPDCAITVHTQADTSS 122
 QY 121 PDPLFNPLKTVGCQLDNANVTDAILERAGSIAFTGHYQTAFRELERVLNFPQSNLCLK 180
 Db 123 PDPLFNPLKTVGCQLDNANVTDAILERAGSIAFTGHYQTAFRELERVLNFPQSNLCLK 182
 QY 181 REKODESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEFGGRITDS 240
 Db 183 REKODESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEFGGRITDS 242
 QY 241 HQMNTLLSLHNAQFDLLIORTPEVARSRATPLDLIKTALTTPHPKQAYGVTLPSTVLFI 300
 Db 243 HQMNTLLSLHNAQFDLLIORTPEVARSRATPLDLIKTALTTPHPKQAYGVTLPSTVLFI 302
 QY 301 AGHDTNLANLGCALELNWTLFGQPDNTPPGGELVFERWRLSDNSQNTQVSLVFTQLQOM 360
 Db 303 AGHDTNLANLGCALELNWTLFGQPDNTPPGGELVFERWRLSDNSQNTQVSLVFTQLQOM 362
 QY 361 RDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 410
 Db 363 RDKTPLSLNTPPGEVKLTLAGCEERNAQGMCSLAGFTQIVNEARIPACSL 412
 RESULT 4
 AAE37853
 ID AAE37853 standard; protein; 431 AA.
 XX AC AAE37853;
 XX DT 06-NOV-2003 (first entry)
 XX DE PNOV4054 phytase fusion protein.
 KW Thermotolerant phytase; weight gain; animals fed diet; grain processing;
 KW nutritive value; transformed plant; anabolic; chimeric; maize.
 XX OS Chimeric - Zea mays.
 XX OS Chimeric - Unidentified.
 XX FH Key Location/Qualifiers
 FT Peptide 1..19
 FT /label= Signal_peptide
 FT Protein 20..431
 FT /note= "Mature pNOV4054 phytase fusion protein"
 XX W02003057248-A1.
 XX 17-JUL-2003.
 XX 30-DEC-2002; 2002WO-US041787.
 XX 28-DEC-2001; 2001US-0344476P.
 XX (SYGN) SYNGENTA PARTICIPATIONS AG.
 XX Lanahan MB, Betts S;
 XX WPI; 2003-607980/57.
 XX N-PSDB; AAD57148.
 XX Preparing a thermotolerant phytase for preparing animal feed or human food by expressing in a plant cell an expression cassette comprising a promoter operably linked to a nucleic acid molecule encoding a thermotolerant phytase.
 XX Claim 63; Page 90; 157pp; English.
 XX The invention relates to a method for preparing a thermotolerant phytase. The method comprises expressing in a plant cell an expression cassette comprising a promoter operably linked to a nucleic acid molecule encoding a thermotolerant phytase which retains at least 40% activity after 30 minutes at 60plusOC and has a specific activity of greater than 200 U/mg at pH 4.5 and 37plusOC. The method is useful for preparing a

thermotolerant phytase for preparing animal feed or human food. The invention is useful for reducing the feed conversion ratio and increasing weight gain, improving reducing feed conversion ratios or increasing weight gain of animals fed diets with inorganic phosphate at levels below 0.4%, minimising dietary requirements of phosphorus in an animal, enhancing the utilisation of phosphorus present in animal feed, enhancing organic phosphorus utilisation from organic phosphorus sources, in animal feed, decreasing the phosphate levels in excreta from an animal, improving the processing of grain, improving the nutritive value of processed grain product or a method of processing grain, improving the nutritive value of animal feed and human food, and preparing a transformed plant which expresses a thermotolerant phytase. The present sequence is PNOV4054 phytase fusion protein used in the exemplification of the invention

XX SQ Sequence 431 AA;

Query Match 100.0%; Score 2153; DB 6; Length 431;
 Best Local Similarity 100.0%; Pred. No. 1.5e-210;
 Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QSEPELKLESVIVSRHGVRAPTKATQLMQDVTDPDAPWTPVKLGELTPRGGELIAYLGH 60
 DB 22 QSEPELKLESVIVSRHGVRAPTKATQLMQDVTDPDAPWTPVKLGELTPRGGELIAYLGH 81
 QY 61 YWQRLVADGLLPGCGPQSGQVAIADVDERTKGTGEAFAGLADPCAITVHTQADTSS 120
 DB 82 YWQRLVADGLLPGCGPQSGQVAIADVDERTKGTGEAFAGLADPCAITVHTQADTSS 141
 QY 121 PDPLFNPLKTVGCOLDNANVTDAILERAGGSIADFTGHYQTAFRELERVLPFQSNCLK 180
 DB 142 PDPLFNPLKTVGCOLDNANVTDAILERAGGSIADFTGHYQTAFRELERVLPFQSNCLK 201
 QY 181 REKQDSCSLTQALPSELKVSADCVSLTGVASLSMLTEIFLLQQAQMPGPGWGRITDS 240
 DB 202 REKQDSCSLTQALPSELKVSADCVSLTGVASLSMLTEIFLLQQAQMPGPGWGRITDS 261
 QY 241 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLDLIKTALTPHPPOKQAYGVTLPSTVLFI 300
 DB 262 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLDLIKTALTPHPPOKQAYGVTLPSTVLFI 321
 QY 301 AGHDTNLNLGGALELNWTLPGQPDNTPPGGELVFERWRLSDNSQWIOVSLVFQTLQOM 360
 DB 322 AGHDTNLNLGGALELNWTLPGQPDNTPPGGELVFERWRLSDNSQWIOVSLVFQTLQOM 381
 QY 361 RDKTPLSLNTPPGEVKLTLAGCBERNAQGMCSLAGFTQIVNEARIPACSL 410
 DB 382 RDKTPLSLNTPPGEVKLTLAGCBERNAQGMCSLAGFTQIVNEARIPACSL 431

RESULT 5

ADC87743
 ID ADC87743 standard; protein; 436 AA.

AC ADC87743;

XX 01-JAN-2004 (first entry)

DE Modified Escherichia coli phytase.

XX Phytase; food supplement; enzyme delivery matrix; soybean meal;
 KW thermotolerance; thermostability; kernel; phytate;
 KW myo-inositol-hexaphosphate; inositol; inorganic phosphate;
 KW thermotolerant; feed value; digestion; enzyme.

XX Synthetic.

OS Escherichia coli.

XX US2003103958-A1.

PN 05-JUN-2003.

XX 24-MAY-2002; 2002US-00156660.

XX 13-AUG-1997; 97US-00910798.
 PR 01-MAR-1999; 99US-00259214.
 PR 13-APR-1999; 99US-00291931.
 PR 25-MAY-1999; 99US-00318528.
 PR 25-MAY-2000; 2000US-00580515.
 PR 24-MAY-2001; 2001US-00866379.
 XX (DIVE-) DIVERSA CORP.
 PA Short JM, Kretz K, Gray KA, Barton NR, Garrett JB, O'donoghue E;
 PI Mathur EJ;
 XX WPI: 2003-787039/74.
 DR N-PSDB; ADC87742.
 XX
 PT New nucleic acid encoding a polypeptide having phytase activity, useful
 PT in improving the feeding value of phytate rich ingredients or as an aid
 PT in phytate digestion.
 XX
 PS Claim 60; SEQ ID NO 2; 113pp; English.
 XX
 CC The invention discloses a new isolated or recombinant nucleic acid which
 CC encodes a polypeptide having a phytase activity. Also claimed is a
 CC nucleic acid probe, an amplification primer sequence pair, an expression
 CC cassette comprising the nucleic acid, a vector comprising the nucleic
 CC acid, a transgenic non-human animal or plant, or its seed, comprising the
 CC nucleic acid, an antisense oligonucleotide, inhibiting the translation of
 CC a phytase message in a cell, a heterodimer comprising the polypeptide and
 CC a second domain, an array comprising immobilised polypeptide or nucleic
 CC acid, a hybridoma comprising an antibody that specifically binds to the
 CC polypeptide, a food supplement for an animal, an edible enzyme delivery
 CC matrix, an edible pellet comprising a granule carrier and the
 CC polypeptide, a feed composition, a soybean meal, isolating or identifying
 CC the polypeptide, making an anti-phytase antibody, producing a recombinant
 CC polypeptide, determining whether a compound binds to the polypeptide,
 CC identifying a modulator, whole cell engineering of new or modified
 CC phenotypes by using real-time metabolic flux analysis, increasing
 CC thermotolerance or thermostability of the phytase polypeptide, increasing
 CC the resistance of the phytase polypeptide to enzymatic inactivation in a
 CC digestive system and processing of corn and sorghum kernels. The phytase
 CC activity comprises catalysis of phytate (myo-inositol-hexaphosphate) to
 CC inositol and inorganic phosphate and is thermotolerant. The nucleic acid
 CC is useful in improving the feeding value of phytate rich ingredients or
 CC as an aid in phytate digestion. The sequence presented is the modified
 CC Escherichia coli phytase.
 XX
 SQ Sequence 436 AA;

Query Match 100.0%; Score 2153; DB 7; Length 436;
 Best Local Similarity 100.0%; Pred. No. 1.6e-210;
 Matches 410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QSEPELKLESVIVSRHGVRAPTKATQLMQDVTDPDAPWTPVKLGELTPRGGELIAYLGH 60
 DB 23 QSEPELKLESVIVSRHGVRAPTKATQLMQDVTDPDAPWTPVKLGELTPRGGELIAYLGH 82
 QY 61 YWQRLVADGLLPGCGPQSGQVAIADVDERTKGTGEAFAGLADPCAITVHTQADTSS 120
 DB 83 YWQRLVADGLLPGCGPQSGQVAIADVDERTKGTGEAFAGLADPCAITVHTQADTSS 142
 QY 121 PDPLFNPLKTVGCOLDNANVTDAILERAGGSIADFTGHYQTAFRELERVLPFQSNCLK 180
 DB 143 PDPLFNPLKTVGCOLDNANVTDAILERAGGSIADFTGHYQTAFRELERVLPFQSNCLK 202
 QY 181 REKQDSCSLTQALPSELKVSADCVSLTGVASLSMLTEIFLLQQAQMPGPGWGRITDS 240
 DB 203 REKQDSCSLTQALPSELKVSADCVSLTGVASLSMLTEIFLLQQAQMPGPGWGRITDS 262
 QY 241 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLDLIKTALTPHPPOKQAYGVTLPSTVLFI 300
 DB 263 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLDLIKTALTPHPPOKQAYGVTLPSTVLFI 322

QY 301 AGHDTNLANLGGALELNWTLPGQPDNTPPGGELVFERWRRLSDNSQWIVSLVFTQLOQM 360
 DB 323 AGHDTNLANLGGALELNWTLPGQPDNTPPGGELVFERWRRLSDNSQWIVSLVFTQLOQM 382
 QY 361 RDKTPLSLNTPPGGVKLTLAGCERNAQGMCSLAGFTQIVNEARIPACSL 410
 DB 383 RDKTPLSLNTPPGGVKLTLAGCERNAQGMCSLAGFTQIVNEARIPACSL 432

RESULT 6
 AAE37854
 ID AAE37854 standard; protein; 437 AA.
 AC AAE37854;
 XX
 DT 06-NOV-2003 (first entry)
 XX
 DE pNOV4058 phytase fusion protein.
 XX
 KW Thermotolerant phytase; weight gain; animals fed diet; grain processing;
 KW nutritive value; transformed plant; anabolic; chimeric; maize.
 OS Chimeric - Zea mays.
 OS Chimeric - Unidentified.
 XX
 XX Key Location/Qualifiers
 FT Peptide 1..119
 FT /label= Signal_peptide
 FT Protein 20..437
 FT /note= "Mature pNOV4058 phytase fusion protein"
 FT Region 432..437
 FT /note= "SEKDEL signal"
 XX
 PN WO2003057248-A1.
 XX
 XX 17-JUL-2003.
 XX
 XX 30-DEC-2002; 2002WO-US041787.
 XX
 XX 28-DEC-2001; 2001US-0344476P.
 XX
 XX (SYGN) SYNGENTA PARTICIPATIONS AG.
 XX
 XX Lanahan ME, Betts S;
 XX
 XX WPI; 2003-607980/57.
 XX
 XX N-PSDB; AAD57149.
 XX
 XX Preparing a thermotolerant phytase for preparing animal feed or human
 XX food by expressing in a plant cell an expression cassette comprising a
 XX promoter operably linked to a nucleic acid molecule encoding a
 XX thermotolerant phytase.
 XX
 XX Example 1; Page 91; 157pp; English.
 XX
 XX The invention relates to a method for preparing a thermotolerant phytase.
 XX The method comprises expressing in a plant cell an expression cassette
 XX comprising a promoter operably linked to a nucleic acid molecule encoding
 XX a thermotolerant phytase which retains at least 40% activity after 30
 XX minutes at 60plusC and has a specific activity of greater than 200 U/mg
 XX at pH 4.5 and 37plusC. The method is useful for preparing a
 XX thermotolerant phytase for preparing animal feed or human food. The
 XX invention is useful for reducing the feed conversion ratio and increasing
 XX weight gain, improving reducing feed conversion ratios or increasing
 XX weight gain of animals fed diets with inorganic phosphate at levels below
 XX 0.45%, minimising dietary requirements of phosphorus in an animal,
 XX enhancing the utilisation of phosphorus present in animal feed, enhancing
 XX organic phosphorus utilisation from organic phosphorus sources in animal
 XX feed, decreasing the phosphate levels in excreta from an animal,
 XX improving the processing of grain, improving the nutritive value of
 XX processed grain product or a method of processing grain, improving the
 XX nutritive value of animal feed and human food, and preparing a
 XX transformed plant which expresses a thermotolerant phytase. The present

CC sequence is pNOV4058 phytase fusion protein used in the exemplification
 CC of the invention
 XX
 SQ Sequence 437 AA;
 Query Match 100.0%; Score 2153; DB 6; Length 437;
 Best Local Similarity 100.0%; Pred. No. 1.6e-210; Indels 0; Gaps 0;
 Matches 410; Conservative 0; Mismatches 0;
 QY 1 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTPDAMPTWPKLGELTPRGGLIAYLGH 60
 DB 22 QSEPELKLESVVIVSRHGVRAPTKATQLMQDVTPDAMPTWPKLGELTPRGGLIAYLGH 81
 QY 61 YNRQELVADGLLPKCGCPQSGOVATIIADVDERTRKTGEAFAAGLAPDCAITVHTQADTSS 120
 DB 82 YNRQELVADGLLPKCGCPQSGOVATIIADVDERTRKTGEAFAAGLAPDCAITVHTQADTSS 141
 QY 121 PDPLEFNPLKTVGCQLDNANVTDAILERAGGSIADFTGHYQTAFARELERVLNPPQSNLCLK 180
 DB 142 PDPLEFNPLKTVGCQLDNANVTDAILERAGGSIADFTGHYQTAFARELERVLNPPQSNLCLK 201
 QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 240
 DB 202 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 261
 QY 241 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLLDLIKLTALTTPHPQKQAYGVTLPSTVLFI 300
 DB 262 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLLDLIKLTALTTPHPQKQAYGVTLPSTVLFI 321
 QY 301 AGHDTNLANLGGALELNWTLPGQPDNTPPGGELVFERWRRLSDNSQWIVSLVFTQLOQM 360
 DB 322 AGHDTNLANLGGALELNWTLPGQPDNTPPGGELVFERWRRLSDNSQWIVSLVFTQLOQM 381
 QY 361 RDKTPLSLNTPPGGVKLTLAGCERNAQGMCSLAGFTQIVNEARIPACSL 410
 DB 382 RDKTPLSLNTPPGGVKLTLAGCERNAQGMCSLAGFTQIVNEARIPACSL 431

RESULT 7
 AAE15808
 ID AAE15808 standard; protein; 430 AA.
 AC AAE15808;
 XX
 XX 26-MAR-2002 (first entry)
 XX
 XX Escherichia coli appA phytase mutant protein.
 XX
 XX Bacterial phytase; K12 appA phytase; protease stability; anabolic;
 XX gastrointestinal; nutritional value; feed treatment process; therapy;
 XX thermal tolerance; growth performance; alcoholic drink; biopulping;
 XX non-alcoholic drink; bioleaching; mutant; muten.
 XX
 XX Escherichia coli.
 OS Synthetic.
 XX
 XX Key Location/Qualifiers
 FH Key
 FT Misc-difference 68 /note= "Wild type Trp substituted with Glu"
 FT Misc-difference 84 /note= "Wild type Gln substituted with Trp"
 FT Misc-difference 95 /note= "Wild type Ala substituted with Pro"
 FT Misc-difference 97 /note= "Wild type Lys substituted with Cys"
 FT Misc-difference 168 /note= "Wild type Ser substituted with Glu"
 FT Misc-difference 180 /note= "Wild type Arg substituted with Tyr"
 FT Misc-difference 225 /note= "Wild type Asn substituted with Cys"
 FT Misc-difference 276 /note= "Wild type Tyr substituted with Asp"

XX WO200190333-A2.
PN 29-NOV-2001.
XX 24-MAY-2001; 2001WO-US017118.
PF 25-MAY-2000; 2000US-00580515.
XX (DIVE-) DIVERSA CORP.
PA Short JM, Kretz KA, Gray KA, Barton NR, Garrett JB, O'donoghue E;
PI WPI; 2002-083108/11.
XX New bacterial phytase for e.g. improving the nutritional value of phytate
PT containing foodstuffs and subsequently improving the growth performance
PT of an organism that consumes it, or in treating animal digestive systems.
XX Claim 59; Fig 8; 170pp; English.
XX The patent discloses recombinant bacterial phytase from Escherichia coli
CC K12 appA phytase. The enzyme has phytase activity and improved thermal
CC tolerance when compared with wild-type phytase. It has improved protease
CC stability at low pH. The recombinant phytase is useful for improving the
CC nutritional value of phytate-containing foodstuffs and subsequently
CC improving the growth performance of an organism that consumes it, in
CC treating animal digestive systems, in feed treatment processes and for in
CC vitro purposes related to research, discovery and development. They are
CC also used for generating recombinant digestive system life forms, for
CC the use of moulds, grains and/or plants, in biopulping and bio-bleaching
CC where a reduction in the use of environmentally harmful chemicals that
CC are traditionally used in the pulp and paper industry is desired and in
CC the reduction or possible elimination of the need for mineral
CC supplements, enzymes or therapeutic drugs for animals from the daily feed
CC thus increasing the amount calories and nutrients present in the feed.
CC The present sequence is E. coli appA phytase mutant protein
XX
XX Sequence 430 AA;
Query Match 98.6%; Score 2123; DB 5; Length 430;
Best Local Similarity 99.5%; Pred. No. 1.8e-207;
Matches 408; Conservative 0; Mismatches 0; Indels 2; Gaps 2;
QY 1 QSEPELKLESVIVSRHGVRAPTKATQMQDVTDPDAPFTWPKLIGELTPRGGLIAYLGH 60
Db 23 QSEPELKLESVIVSRHGVRAPTKATQMQDVTDPDAPFTWPKLIGELTPRGGLIAYLGH 82
QY 61 YWRQLVADGLLPKCGCPQSQVAILIADVDRTRKTGEAFAGLAPDCAITVHTQADTSS 120
Db 83 YWRQLVADGLLPKCGCPQSQVAILIADVDRTRKTGEAFAGLAPDCAITVHTQADTSS 142
QY 121 PDPLFNPLKTVGCOLDNANVTDAILERAGGSIAFTGHYQATPRELERNVFPQSNICLK 180
Db 143 PDPLFNPLKTVGCOLDNANVTDAILE-AGGSIAFTGHYQATPRELERNVFPQSNICLK 201
QY 181 REKQDESCLTQALPSELKVSADCVSLTGVASLSASMLTEIFLQQAQMPFGHGRITDS 240
Db 202 REKQDESCLTQALPSELKVSADCVSLTGVASLSASMLTEIFLQQAQMPFGHGRITDS 261
QY 241 HOWNTLISHNAQFLLORTTEVARSRATPLDLIKTALTPHPQKQAYGVTLTSLVFI 300
Db 262 HOWNTLISHNAQFLLORTTEVARSRATPLDLIKTALTPHPQKQAYGVTLTSLVFI 320
QY 301 AGHDTNLANLGALBELNLTLPQDPNTPPGGELVFERWRRLSDNSQWITQVSLVFTQLOOM 360
Db 321 AGHDTNLANLGALBELNLTLPQDPNTPPGGELVFERWRRLSDNSQWITQVSLVFTQLOOM 380
QY 361 RDKTPLSLNTPPGEVKLTLAGCEERNAQCMCSLAGFTQIVNEARIPACSL 410
Db 381 RDKTPLSLNTPPGEVKLTLAGCEERNAQCMCSLAGFTQIVNEARIPACSL 430

RESULT 8
ADA19451
ID ADA19451 standard; protein; 430 AA.
XX
XX ADA19451;
XX
XX 20-NOV-2003 (first entry)
XX
XX E. coli K12 phytase mutant.
XX
XX Phytase; enzyme; phytate; appA gene; animal feed; inorganic phosphate;
KW digestion enhancement; transgenic; thermal tolerance; protease stability;
KW mutant; mutein.
XX
XX Synthetic.
OS Escherichia coli; strain K12.
OS
XX
XX Key Location/Qualifiers
FH Misc-difference 68 /note= "Wild-type Trp substituted by Glu"
FT Misc-difference 72 /note= "Encoded by CGN"
FT Misc-difference 84 /note= "Wild-type Gln substituted by Trp"
FT Misc-difference 95 /note= "Wild-type Ala substituted by Pro"
FT Misc-difference 97 /note= "Wild-type Lys substituted by Cys"
FT Misc-difference 168 /note= "Wild-type Ser substituted by Glu"
FT Misc-difference 180 /note= "Wild-type Arg substituted by Tyr"
FT Misc-difference 226 /note= "Wild-type Asn substituted by Cys"
FT Misc-difference 277 /note= "Wild-type Tyr substituted by Asp"
XX
XX US2002136754-A1.
XX
XX 26-SEP-2002.
XX
XX 24-MAY-2001; 2001US-00866379.
XX
XX 13-AUG-1997; 97US-00910798.
XX 01-MAR-1999; 99US-00259214.
XX 13-APR-1999; 99US-00231931.
XX 25-MAY-1999; 99US-00318528.
XX 25-MAY-2000; 2000US-00580515.
XX
XX (SHOR/) SHORT J M.
PA (KRET/) KRETZ K A.
PA (GRAY/) GRAY K A.
PA (BART/) BARTON N R.
PA (GARE/) GARRETT J B.
PA (DONO/) O' DONOGHUE E.
PA (MATH/) MATHUR E J.
XX
XX Short JM, Kretz KA, Gray KA, Barton NR, Garrett JB;
PI O' Donoghue E, Mathur EJ;
XX WPI; 2003-040002/03.
XX
XX Isolated Escherichia coli polynucleotide encoding a modified phytase
PT enzyme, useful in the production of animal feed, for improving the
PT nutritional value of phytate-containing foodstuff and for enhancing
PT digestion in humans and animals.
XX
XX Claim 59; Fig 8; 62pp; English.
XX
XX The invention relates to an isolated Escherichia coli polynucleotide
CC encoding a phytase enzyme appearing as ADA19450 and having amino acids
CC modification from W68E, Q84W, A95P, K97C, S168E, R180Y, N226C or Y277D.

CC Also included the E. coli appA gene ADA19449 (or an oligonucleotide
CC derived from it) or its mutant sequence ADA19452, expression vectors,
CC host cells, a method of improving nutritional value of a phytate-
CC containing foodstuff by contacting the phytate-containing foodstuff with
CC a pure phytase enzyme mutated phytase (where the phytase enzyme catalyzes
CC the liberation of inorganic phosphate from the phytate in the phytate-
CC containing foodstuff), a method to produce an animal feed containing a
CC microbial phytase (comprising culturing the plant cell, plant part or
CC plant under conditions where the nucleotide sequence is expressed and
CC converting the plant cells, plants or plant into a composition for animal
CC feed), a feed composition for animals (comprising the plant seeds, plant
CC cells, plant parts or plants in admixture with a phytate-containing
CC foodstuff), a method to treat a human or an animal able to benefit from
CC digestive enhancement by the activity of an exogenous phytase enzymes
CC comprising administering to the human or animal the plant seed, plant
CC cells, plant parts or plants of a transgenic plant which is modified to
CC contain an expression system which expresses a nucleotide sequence
CC encoding a phytase enzyme, a transgenic non-human organism whose genome
CC comprising a heterologous nucleic acid sequence encoding a polypeptide
CC having phytase activity. The phytase enzyme is useful for improving the
CC nutritional value of phytate-containing foodstuff, in the production of
CC animal feed and for enhancing digestion in humans and animals. The
CC invented method improves thermal tolerance and protease stability. It
CC also improves the feeding value of phytate rich ingredients. The present
CC sequence represents E. coli K12 mutant phytase.

XX Sequence 430 AA;

Query Match 98.6%; Score 2123; DB 6; Length 430;
Best Local Similarity 99.5%; Pred. No. 1.8e-207;
Matches 408; Conservative 0; Mismatches 0; Indels 2; Gaps 2;
QY 1 QSEPELKESVIVSRHGVRAPTKATQLMQDVTDPANFTWPKLGTTPRGGELIAYLGH 60
DB 23 QSEPELKESVIVSRHGVRAPTKATQLMQDVTDPANFTWPKLGTTPRGGELIAYLGH 82
QY 61 YRQRLVADGLPKCGCPQSGQVAILADVDETRTKTGEAFAGLAPDCAITVHTQADTSS 120
DB 83 YRQRLVADGLPKCGCPQSGQVAILADVDETRTKTGEAFAGLAPDCAITVHTQADTSS 142
QY 121 PDPLENPLKTVGCQLDNANVTDAILERAGGSIADFTGHYQTAFARELERVLFNFPQSNCLK 180
DB 143 PDPLENPLKTVGCQLDNANVTDAILE-AGGSIADFTGHYQTAFARELERVLFNFPQSNCLK 201
QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLOQAQMPGPGGRITDS 240
DB 202 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLOQAQMPGPGGRITDS 261
QY 241 HQWNTLLSLHNAQFDLQRTPEVARSRATPLDLIKTALTALTPHPKQAYGVTLFTSVLFI 300
DB 262 HQWNTLLSLHNAQFD-LQRTPEVARSRATPLDLIKTALTALTPHPKQAYGVTLFTSVLFI 320
QY 301 AGHDTNLANLGALSLNWTLPQDPNTPPGGELVFERWRRLSDNSQWISVLFQTLQOM 360
DB 321 AGHDTNLANLGALSLNWTLPQDPNTPPGGELVFERWRRLSDNSQWISVLFQTLQOM 380
QY 361 RDKTPLSLNTPPGGVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 410
DB 381 RDKTPLSLNTPPGGVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 430

RESULT 9

AAB36257
ID AAB36257 standard; protein; 432 AA.
XX
AC AAB36257;
XX
XX
DT 12-SEP-2003 (revised)
DT 20-FEB-2001 (first entry)
XX
XX
DE Lama2/APPA plasmid translated sequence.
XX Transgenic animal; salivary protein; phytase; phosphorus; animal growth;

XX environmental pollution; pig.
XX Mus musculus.
OS Escherichia coli.
OS Chimeric.
XX WO200064247-A1.
FN 02-NOV-2000.
PD 20-APR-2000; 2000WO-CA000430.
PF 23-APR-1999; 99US-0130508P.
PR (UYGU-) UNIV GUELPH.
XX Forsberg CW, Golovan S, Phillips JP;
PI WPI; 2000-687245/67.
DR N-PSDB; AAC68294.
XX Transgenic non-human animal for gastrointestinal tract specific
PT expression of a protein, preferably phytase, comprises a nucleic acid
PT sequence including a heterologous transgene construct encoding the
PT protein.
XX Disclosure; Fig 5; 152pp; English.

XX The present invention provides transgenic animals which produce desired
CC proteins, in this case pigs which expresses phytase in the salivary
CC gland. Low phytase production levels result in phytate in the diet being
CC excreted and causing phosphorus contamination in water, as well as
CC reducing the growth of animals. The invention provides a number of
CC transgenes containing the E. coli APPA phytase coding sequence. (Updated
CC on 12-SEP-2003 to standardise OS field)

XX Sequence 432 AA;

Query Match 96.5%; Score 2077; DB 3; Length 432;
Best Local Similarity 98.0%; Pred. No. 9e-203;
Matches 402; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 1 QSEPELKESVIVSRHGVRAPTKATQLMQDVTDPANFTWPKLGTTPRGGELIAYLGH 60
DB 23 QSEPELKESVIVSRHGVRAPTKATQLMQDVTDPANFTWPKLGTTPRGGELIAYLGH 82
QY 61 YRQRLVADGLPKCGCPQSGQVAILADVDETRTKTGEAFAGLAPDCAITVHTQADTSS 120
DB 83 YRQRLVADGLPKCGCPQSGQVAILADVDETRTKTGEAFAGLAPDCAITVHTQADTSS 142
QY 121 PDPLENPLKTVGCQLDNANVTDAILERAGGSIADFTGHYQTAFARELERVLFNFPQSNCLK 180
DB 143 PDPLENPLKTVGCQLDNANVTDAILESRAGSIADFTGHYQTAFARELERVLFNFPQSNCLK 202
QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLOQAQMPGPGGRITDS 240
DB 203 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLOQAQMPGPGGRITDS 262
QY 241 HQWNTLLSLHNAQFDLQRTPEVARSRATPLDLIKTALTALTPHPKQAYGVTLFTSVLFI 300
DB 263 HQWNTLLSLHNAQFDLQRTPEVARSRATPLDLIKTALTALTPHPKQAYGVTLFTSVLFI 322
QY 301 AGHDTNLANLGALSLNWTLPQDPNTPPGGELVFERWRRLSDNSQWISVLFQTLQOM 360
DB 323 AGHDTNLANLGALSLNWTLPQDPNTPPGGELVFERWRRLSDNSQWISVLFQTLQOM 382
QY 361 RDKTPLSLNTPPGGVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 410
DB 383 RDKTPLSLNTPPGGVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 432

RESULT 10
AAB36261

ID XX AC AAB36261; 323 AGHDTNLANLGALSNLWTLPGQPDNTPPGGELVFERWRRRLSDNSQWISLVFTQLQQM 382

AC XX AAB36262; 361 RDKTFLSLNTPPGEVKLTLAGCEERNAQMGSLAGFTQIVNEARIPACSL 410

DT XX 12-SEP-2003 (revised) 383 RDKTFLSLNTPPGEVKLTLAGCEERNAQMGSLAGFTQIVNEARIPACSL 432

DT XX 20-FEB-2001 (first entry)

DE XX R15/APPA plasmid translated sequence.

XX KW Transgenic animal; salivary protein; phytase; phosphorus; animal growth;

XX KW environmental pollution; pig.

OS Rattus sp.

OS Escherichia coli.

OS Chimeric.

XX WO200064247-A1.

XX 02-NOV-2000.

XX 20-APR-2000; 2000WO-CA000430.

XX 23-APR-1999; 99US-0130508P.

XX (UYGU-) UNIV GUELPH.

XX Forsberg CW, Golovan S, Phillips JP;

XX WPI; 2000-687245/67.

XX N-PSDB; AAC68298.

XX Transgenic non-human animal for gastrointestinal tract specific

XX expression of a protein, preferably phytase, comprises a nucleic acid

XX sequence including a heterologous transgene construct encoding the

XX protein.

XX Disclosure; Fig 21; 152pp; English.

XX The present invention provides transgenic animals which produce desired

XX proteins, in this case pigs which expresses phytase in the salivary

XX gland. Low phytase production levels result in phytate in the diet being

XX excreted and causing phosphorus contamination in water, as well as

XX reducing the growth of animals. The invention provides a number of

XX transgenes containing the E. coli APPA phytase coding sequence. (Updated

XX on 12-SEP-2003 to standardise OS field)

XX Sequence 432 AA;

Query Match 96.5%; Score 2077; DB 3; Length 432;

Best Local Similarity 98.0%; Pred. No. 9e-203;

Matches 402; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 QSEPELKLESVVVSRHGVRAFTKATQMQDVTDPAMPTWPKLGLTPRGGLIAYLGH 60

DB 23 QSEPELKLESVVVSRHGVRAFTKATQMQDVTDPAMPTWPKLGLTPRGGLIAYLGH 82

QY 61 YRQRLVADGLLPKCGCPQSGQVIAIDVDETRTKTGEAFAGLAPDCAITVHTQADTSS 120

DB 83 YRQRLVADGLLPKCGCPQSGQVIAIDVDETRTKTGEAFAGLAPDCAITVHTQADTSS 142

QY 121 PDPLFNPLKTGVCQDLNANVTDAILERAGGSIADFTGHYQTAFAFRELVRNFPQSNCLK 180

DB 143 PDPLFNPLKTGVCQDLNANVTDAILERAGGSIADFTGHYQTAFAFRELVRNFPQSNCLK 202

QY 181 REKQDESCLTQALPSELKVSADCVSLTGAVSLASMLTEIFLQOAGMPFGHGRITDS 240

DB 203 REKQDESCLTQALPSELKVSADCVSLTGAVSLASMLTEIFLQOAGMPFGHGRITDS 262

QY 241 HOWNTLLSHNAQFOLLORTTEVARSRATPLDLIKTALTTPHPQKQAYGVTLTSLVFI 300

DB 263 HOWNTLLSHNAQFOLLORTTEVARSRATPLDLIKTALTTPHPQKQAYGVTLTSLVFI 322

QY 301 AGHDTNLANLGALSNLWTLPGQPDNTPPGGELVFERWRRRLSDNSQWISLVFTQLQQM 360

DB 323 AGHDTNLANLGALSNLWTLPGQPDNTPPGGELVFERWRRRLSDNSQWISLVFTQLQQM 382

QY 361 RDKTFLSLNTPPGEVKLTLAGCEERNAQMGSLAGFTQIVNEARIPACSL 410

DB 383 RDKTFLSLNTPPGEVKLTLAGCEERNAQMGSLAGFTQIVNEARIPACSL 432

RESULT 11

AAB36262

ID AAB36262 standard; protein; 432 AA.

XX AC AAB36262;

XX 12-SEP-2003 (revised)

DT 20-FEB-2001 (first entry)

XX SV40/APPA plasmid translated sequence.

XX KW Transgenic animal; salivary protein; phytase; phosphorus; animal growth;

XX KW environmental pollution; pig.

XX OS Simian virus 40.

OS Escherichia coli.

OS Chimeric.

XX WO200064247-A1.

XX 02-NOV-2000.

XX 20-APR-2000; 2000WO-CA000430.

XX 23-APR-1999; 99US-0130508P.

XX (UYGU-) UNIV GUELPH.

XX Forsberg CW, Golovan S, Phillips JP;

XX WPI; 2000-687245/67.

XX N-PSDB; AAC68298.

XX Transgenic non-human animal for gastrointestinal tract specific

XX expression of a protein, preferably phytase, comprises a nucleic acid

XX sequence including a heterologous transgene construct encoding the

XX protein.

XX Disclosure; Fig 22; 152pp; English.

XX The present invention provides transgenic animals which produce desired

XX proteins, in this case pigs which expresses phytase in the salivary

XX gland. Low phytase production levels result in phytate in the diet being

XX excreted and causing phosphorus contamination in water, as well as

XX reducing the growth of animals. The invention provides a number of

XX transgenes containing the E. coli APPA phytase coding sequence. (Updated

XX on 12-SEP-2003 to standardise OS field)

XX Sequence 432 AA;

Query Match 96.5%; Score 2077; DB 3; Length 432;

Best Local Similarity 98.0%; Pred. No. 9e-203;

Matches 402; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 QSEPELKLESVVVSRHGVRAFTKATQMQDVTDPAMPTWPKLGLTPRGGLIAYLGH 60

DB 23 QSEPELKLESVVVSRHGVRAFTKATQMQDVTDPAMPTWPKLGLTPRGGLIAYLGH 82

QY 61 YRQRLVADGLLPKCGCPQSGQVIAIDVDETRTKTGEAFAGLAPDCAITVHTQADTSS 120

DB 83 YRQRLVADGLLPKCGCPQSGQVIAIDVDETRTKTGEAFAGLAPDCAITVHTQADTSS 142

QY 121 PDPLFNPLKTGVCQDLNANVTDAILERAGGSIADFTGHYQTAFAFRELVRNFPQSNCLK 180

DB 143 PDPLFNPLKTGVCQDLNANVTDAILERAGGSIADFTGHYQTAFAFRELVRNFPQSNCLK 202

QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMPGPGHGRITDS 240
DB 203 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMPGPGHGRITDS 262
QY 241 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLDLDIKTALTTEHPPOKQAYGVLTPTSVLFI 300
DB 263 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLDLDIKTALTTEHPPOKQAYGVLTPTSVLFI 322
QY 301 AGHDTNLANLGGALELNTLPGQPDNTPPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 360
DB 323 AGHDTNLANLGGALELNTLPGQPDNTPPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 382
QY 361 RDKTPLSLNTPPGGEVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 410
DB 383 RDKTPLSLNTPPGGEVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 432

RESULT 12
ID AAB36259 standard; protein; 432 AA.
AC AAB36259;
XX
DT 12-SEP-2003 (revised)
DT 20-FEB-2001 (first entry)
XX
DE R15/APPA plasmid translated sequence.
XX
XX Transgenic animal; salivary protein; phytase; phosphorus; animal growth;
XX environmental pollution; pig.
OS Rattus sp.
OS Escherichia coli.
OS Chimeric.
XX
XX WO200064247-A1.
XX
XX 02-NOV-2000.
XX
XX 20-APR-2000; 2000WO-CA000430.
XX
XX 23-APR-1999; 99US-0130508P.
XX
XX (UYGU-) UNIV GUELPH.
XX
XX Forsberg CW, Golovan S, Phillips JP;
XX
XX WPI; 2000-687245/67.
XX
XX N-PSDB; AAC68296.
XX
XX Transgenic non-human animal for gastrointestinal tract specific
XX expression of a protein, preferably phytase, comprises a nucleic acid
XX sequence including a heterologous transgene construct encoding the
XX protein.
XX
XX Disclosure; Fig 19; 152pp; English.
XX
XX The present invention provides transgenic animals which produce desired
XX proteins, in this case pigs which expresses phytase in the salivary
XX gland. Low phytase production levels result in phytate in the diet being
XX excreted and causing phosphorus contamination in water, as well as
XX reducing the growth of animals. The invention provides a number of
XX transgenes containing the E. coli APPA phytase coding sequence. (Updated
XX on 12-SEP-2003 to standardise OS field)
XX
XX Sequence 432 AA;
XX
XX Query Match 96.5%; Score 2077; DB 3; Length 432;
XX Best Local Similarity 98.0%; Pred. No. 9a-203;
XX Matches 402; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
XX
XX 1 QSEPELKLESVVIVSRHGVRATKATQMLQMDVTPDANPTFWKLGELTPRGGELIAYLGH 60

DB 23 QSEPELKLESVVIVSRHGVRATKATQMLQMDVTPDANPTFWKLGELTPRGGELIAYLGH 82
QY 61 YNRQRLVADGLLPKCCPGSQGQVAILADVDERTRKTGEFAAGLAPDCAITVHTQADTSS 120
DB 83 YQRQRLVADGLLPKCCPGSQGQVAILADVDERTRKTGEFAAGLAPDCAITVHTQADTSS 142
QY 121 PDLFENPLKTVGCOLDNANVTDAILLERAGGSTADFTGHVQTAFRELERVLPQSNLCLK 180
DB 143 PDLFENPLKTVGCOLDNANVTDAILLERAGGSTADFTGHVQTAFRELERVLPQSNLCLK 202
QY 181 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMPGPGHGRITDS 240
DB 203 REKQDESCSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQMPGPGHGRITDS 262
QY 241 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLDLDIKTALTTEHPPOKQAYGVLTPTSVLFI 300
DB 263 HQWNTLLSLHNAQFDLLQRTPEVARSRATPLDLDIKTALTTEHPPOKQAYGVLTPTSVLFI 322
QY 301 AGHDTNLANLGGALELNTLPGQPDNTPPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 360
DB 323 AGHDTNLANLGGALELNTLPGQPDNTPPGGELVFERWRRLSDNSQWIOVSLVFQTLQOM 382
QY 361 RDKTPLSLNTPPGGEVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 410
DB 383 RDKTPLSLNTPPGGEVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 432

RESULT 13
ID AAB36258 standard; protein; 432 AA.
XX
XX AAB36258;
XX
DT 12-SEP-2003 (revised)
DT 20-FEB-2001 (first entry)
XX
XX R15/APPA plasmid translated sequence.
XX
XX Transgenic animal; salivary protein; phytase; phosphorus; animal growth;
XX environmental pollution; pig.
OS Rattus sp.
OS Escherichia coli.
OS Chimeric.
XX
XX WO200064247-A1.
XX
XX 02-NOV-2000.
XX
XX 20-APR-2000; 2000WO-CA000430.
XX
XX 23-APR-1999; 99US-0130508P.
XX
XX (UYGU-) UNIV GUELPH.
XX
XX Forsberg CW, Golovan S, Phillips JP;
XX
XX WPI; 2000-687245/67.
XX
XX N-PSDB; AAC68295.
XX
XX Transgenic non-human animal for gastrointestinal tract specific
XX expression of a protein, preferably phytase, comprises a nucleic acid
XX sequence including a heterologous transgene construct encoding the
XX protein.
XX
XX Disclosure; Fig 18; 152pp; English.
XX
XX The present invention provides transgenic animals which produce desired
XX proteins, in this case pigs which expresses phytase in the salivary
XX gland. Low phytase production levels result in phytate in the diet being
XX excreted and causing phosphorus contamination in water, as well as
XX reducing the growth of animals. The invention provides a number of

CC transgenes containing the E. coli APPA phytase coding sequence. (Updated
CC on 12-SEP-2003 to standardise OS field)
XX
SQ Sequence 432 AA;

Query Match 96.5%; Score 2077; DB 3; Length 432;
Best Local Similarity 98.0%; Pred. No. 9e-203;
Matches 402; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 QSEPELKLESVVIVSRHGVRAPTKATQMQDVTDPDAMPTWPKLGELTPRGGLIAYLGH 60
DB 23 QSEPELKLESVVIVSRHGVRAPTKATQMQDVTDPDAMPTWPKLGELTPRGGLIAYLGH 82

QY 61 YWRQRLVADGLPKCGCPQSGQVAIIADVDERTRKTGEAFAGLAPDCAITVHTQADTSS 120
DB 83 YORQLVADGLLAKKGCPSQGVALLIADVDERTRKTGEAFAGLAPDCAITVHTQADTSS 142

QY 121 PDPLFNLKTVGCQLDNANVTDAILERAGGSIADFTGHYQTAFRELERVLPNPPQSNLCLK 180
DB 143 PDPLFNLKTVGCQLDNANVTDAILERAGGSIADFTGHYQTAFRELERVLPNPPQSNLCLK 202

QY 181 REKQDECSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQOAGMPEPGWGRITDS 240
DB 203 REKQDECSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQOAGMPEPGWGRITDS 262

QY 241 HQWNTLLSLHNAQFVLLQRTPEVARSRATPLLDLIKLTALTPHPPOKQAYGVTLPSTVLFI 300
DB 263 HQWNTLLSLHNAQFVLLQRTPEVARSRATPLLDLIKLTALTPHPPOKQAYGVTLPSTVLFI 322

QY 301 AGHDTNLANLGGALELNLWTLPGQDNTPPGGELVFERWRRLSDNSQWISLVFOTLQOM 360
DB 323 AGHDTNLANLGGALELNLWTLPGQDNTPPGGELVFERWRRLSDNSQWISLVFOTLQOM 382

QY 361 RDKTPLSLNTPPGVEVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 410
DB 383 RDKTPLSLNTPPGVEVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 432

RESULT 14
AAB36263
ID AAB36263 standard; protein; 432 AA.

AC AAB36263;
XX
XX
DT 12-SEP-2003 (revised)
DT 20-FEB-2001 (first entry)

Lama2/APPA plasmid translated sequence.
Transgenic animal; salivary protein; phytase; phosphorus; animal growth;
environmental pollution; pig.
Mus musculus.
Escherichia coli.
Chimeric.
PN WO200064247-A1.
XX
XX
PD 02-NOV-2000.
XX
PF 20-APR-2000; 2000WO-CA000430.
XX
PR 23-APR-1999; 99US-0130508P.
XX
PA (UYGU-) UNIV GUELPH.
XX
PI Forsberg CW, Golovan S, Phillips JP;
XX
XX WPI; 2000-687245/67.
DR N-PSDB; AAC68300.
XX
PT Transgenic non-human animal for gastrointestinal tract specific
expression of a protein, preferably phytase, comprises a nucleic acid

PT sequence including a heterologous transgene construct encoding the
PT protein.
XX
PS Disclosure; Fig 23; 152pp; English.
XX
CC The present invention provides transgenic animals which produce desired
CC proteins, in this case pigs which expresses phytase in the salivary
CC gland. Low phytase production levels result in phytate in the diet being
CC excreted and causing phosphorus contamination in water, as well as
CC reducing the growth of animals. The invention provides a number of
CC transgenes containing the E. coli APPA phytase coding sequence. (Updated
CC on 12-SEP-2003 to standardise OS field)
XX
SQ Sequence 432 AA;

Query Match 96.5%; Score 2077; DB 3; Length 432;
Best Local Similarity 98.0%; Pred. No. 9e-203;
Matches 402; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 QSEPELKLESVVIVSRHGVRAPTKATQMQDVTDPDAMPTWPKLGELTPRGGLIAYLGH 60
DB 23 QSEPELKLESVVIVSRHGVRAPTKATQMQDVTDPDAMPTWPKLGELTPRGGLIAYLGH 82

QY 61 YWRQRLVADGLPKCGCPQSGQVAIIADVDERTRKTGEAFAGLAPDCAITVHTQADTSS 120
DB 83 YORQLVADGLLAKKGCPSQGVALLIADVDERTRKTGEAFAGLAPDCAITVHTQADTSS 142

QY 121 PDPLFNLKTVGCQLDNANVTDAILERAGGSIADFTGHYQTAFRELERVLPNPPQSNLCLK 180
DB 143 PDPLFNLKTVGCQLDNANVTDAILERAGGSIADFTGHYQTAFRELERVLPNPPQSNLCLK 202

QY 181 REKQDECSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQOAGMPEPGWGRITDS 240
DB 203 REKQDECSLTQALPSELKVSADCVSLTGAVSLASMLTEIFLLQOAGMPEPGWGRITDS 262

QY 241 HQWNTLLSLHNAQFVLLQRTPEVARSRATPLLDLIKLTALTPHPPOKQAYGVTLPSTVLFI 300
DB 263 HQWNTLLSLHNAQFVLLQRTPEVARSRATPLLDLIKLTALTPHPPOKQAYGVTLPSTVLFI 322

QY 301 AGHDTNLANLGGALELNLWTLPGQDNTPPGGELVFERWRRLSDNSQWISLVFOTLQOM 360
DB 323 AGHDTNLANLGGALELNLWTLPGQDNTPPGGELVFERWRRLSDNSQWISLVFOTLQOM 382

QY 361 RDKTPLSLNTPPGVEVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 410
DB 383 RDKTPLSLNTPPGVEVKLTLAGCEERNAQMGCSLAGFTQIVNEARIPACSL 432

RESULT 15
AAB36260
ID AAB36260 standard; protein; 432 AA.

AC AAB36260;
XX
XX
DT 12-SEP-2003 (revised)
DT 20-FEB-2001 (first entry)

R15/APPA plasmid translated sequence.
Transgenic animal; salivary protein; phytase; phosphorus; animal growth;
environmental pollution; pig.
Rattus sp.
Escherichia coli.
Chimeric.
OS
XX
PN WO200064247-A1.
XX
PD 02-NOV-2000.
XX
PF 20-APR-2000; 2000WO-CA000430.
XX
PR 23-APR-1999; 99US-0130508P.
PR

XX (UYGU-) UNIV GUELPH.
XX Forsberg CW, Golovan S, Phillips JP;
XX WPI; 2000-687245/67.
XX DR N-PSDB; AAC68297.
XX Transgenic non-human animal for gastrointestinal tract specific
PT expression of a protein, preferably phytase, comprises a nucleic acid
PT sequence including a heterologous transgene construct encoding the
PT protein.
XX
XX Disclosure; Fig 20; 152pp; English.
XX
XX The present invention provides transgenic animals which produce desired
CC proteins, in this case pigs which expresses phytase in the salivary
CC gland. Low phytase production levels result in phytate in the diet being
CC excreted and causing phosphorus contamination in water, as well as
CC reducing the growth of animals. The invention provides a number of
CC transgenes containing the E. coli APPA phytase coding sequence. (Updated
CC on 12-SEP-2003 to standardise OS field)
XX
XX Sequence 432 AA;
SQ
Query Match 96.5%; Score 2077; DB 3; Length 432;
Best Local Similarity 98.0%; Pred. No. 9e-203; 8; Indels 0; Gaps 0;
Matches 402; Conservative 0; Mismatches 0;
QY 1 QSEPELKLESVTVSRHGVRAPTKATQMQDVTDPDAPWTPVKLGELTPRGGELIAYLGH 60
Db 23 QSEPELKLESVTVSRHGVRAPTKATQMQDVTDPDAPWTPVKLGELIAYLGH 82
QY 61 YWQRIVADGLLPKCCPOSSQGVAIADVDERTKRTGEAFAAGLAPDCAITVHTQADTSS 120
Db 83 YWQRIVADGLLAKGCPQSGQVAIADVDERTKRTGEAFAAGLAPDCAITVHTQADTSS 142
QY 121 PDPLFNPLKTGVCOLDNANVTDAILRAGGSIAADFTGHYQTAFRELERVLPQSNCLIK 180
Db 143 PDPLFNPLKTGVCOLDNANVTDAILRAGGSIAADFTGHYQTAFRELERVLPQSNCLIK 202
QY 181 REKQDESCSLTOALPSELKVSADCVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 240
Db 203 REKQDESCSLTOALPSELKVSADNVSLTGAVSLASMLTEIFLLQQAQGMPEPGWGRITDS 262
QY 241 HQWNTLLSLHNAQFDLQRTPEVARSRATPELLDLIKTALTTPHPQQAQGVTLPTSVLFI 300
Db 263 HQWNTLLSLHNAQFDLQRTPEVARSRATPELLDLIKTALTTPHPQQAQGVTLPTSVLFI 322
QY 301 AGHDTNLNLGGALELNTLPGQPDNTPPGGELVFERWRRLSDNSQWIQVSLVFQTLQOM 360
Db 323 AGHDTNLNLGGALELNTLPGQPDNTPPGGELVFERWRRLSDNSQWIQVSLVFQTLQOM 382
QY 361 RDKTPLSLNTPPGVEVKLTLAGCERNAGQCSLAGFTQIVNEARIACSL 410
Db 383 RDKTPLSLNTPPGVEVKLTLAGCERNAGQCSLAGFTQIVNEARIACSL 432